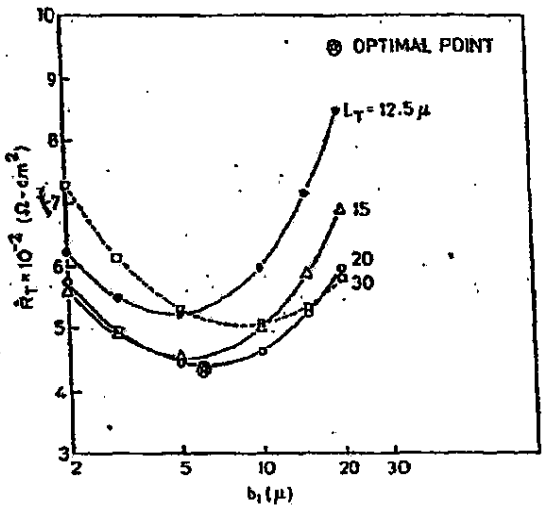


Fig. 2.26. NORMALIZED ON-RESISTANCE VS CELL WIDTH FOR VARIOUS GATE WIDTHS.

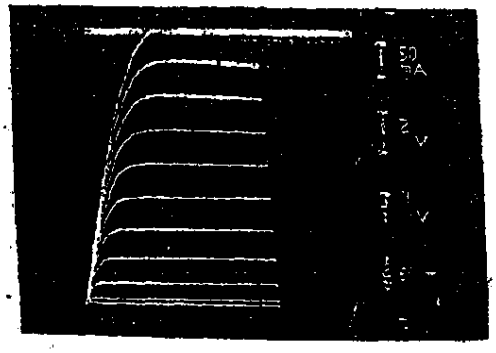
G. Device Transconductance

The low drain-voltage on-resistance of various power MOSFETs is the principal parameter in most switching applications. In linear applications, however, device transconductance g_m in the saturation region is important. This section discusses g_m in the saturation region, and Appendix A will further describe it in the linear region.

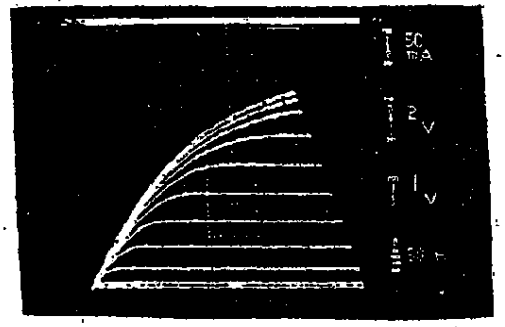
Short-channel transconductance has been well modeled in the past [2.13,2.21] after incorporation of the mobility-reduction factor and velocity-saturation effects. In power MOSFETs, the two major factors that can affect g_m behavior are drift-region resistance and the heat generated from it.

If the device is well into the saturation region, g_m will be independent of the drift region (R_{drift} has no effect on g_m; however, source resistance R_S will be detrimental). This will not be true if the temperature rise caused by I^2R heat generation is considered, which often occurs in high-voltage power MOSFETs.

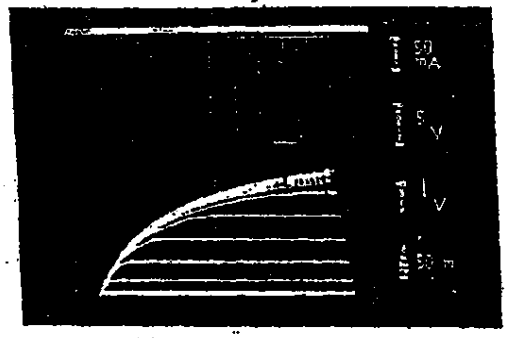
For gate voltages of less than



a. 0.5 Ω-cm



b. 3 Ω-cm



c. 7 Ω-cm

Fig. 2.27. EXPERIMENTAL I-V CHARACTERISTICS OF VDMOS DEVICES FABRICATED WITH VARIOUS EPITAXIAL RESISTIVITIES.

$$\frac{V_{GS} - V_{TE}}{L_{eff}} < E_{crit}$$

where E_{crit} is the critical electric field that causes velocity saturation ($\approx 1.5 \times 10^4$ V/cm), device transconductance is directly related to low-field electron mobility and is approximated [2.30,2.31] by

$$g_m = \frac{W}{L_{eff}} C_o \mu_E (V_{GS} - V_{TE}) \qquad V_{DS} \geq (V_{GS} - V_{TE}) \qquad (2.45)$$

At higher gate voltages and for scattering-limited velocity operation which is typical of these devices in their saturation regions because of their very short channels, transconductance to first order [2.32] is

$$g_{m_{MAX}} = C_o W v_{SAT} \qquad (2.46)$$

where v_{SAT} is the electron scattering-limited velocity (6.5×10^6 cm/sec in <100> silicon inversion layers). Typically, transconductance in these short channels initially increases as $V_{GS} - V_{TE}$ rises, with a slope proportional to μ_E, and gradually reaches the limiting value in Eq. (2.46) as velocity saturation is achieved.

The curve-tracer photographs in Fig. 2.27 illustrate this behavior. The maximum transconductance values of these three VDMOS devices are approximately 50 to 60 mmhos which is somewhat less than the ≈75 mmhos predicted by Eq. (2.46). It is apparent that drift-region bulk resistance limits the current substantially, particularly in the 7 Ω-cm device but also in the 3 Ω-cm structure. (Note the different horizontal scale in Fig. 2.27c.) Much higher drain voltages are required, therefore, to maintain operation in the saturation region as current increases and, as a result, power dissipation and temperature will rise. Because low field mobility and scattering-limited velocity decrease with increasing temperature [2.6,2.7], device gain is also expected to fall. Both the steady-state chip temperature caused by average power dissipation and the "instantaneous" temperature under transient power dissipation are important in determining the effective values for μ_E and v_{SAT} [2.33].

These effects result typically in the g_m vs $V_G - V_{TE}$ experimental curves in Fig. 2.28. These measurements were obtained from three VDMOS and one VMOS device (W = 3440 μ in each), using 80 μsec pulsed measurements on a standard curve tracer. For each gate voltage, the drain voltage was adjusted to assure operation in the saturation region, which normally requires greater drain voltages in the higher epitaxial-resistivity devices. The 0.5 Ω-cm VDMOS achieves a maximum g_m close to that predicted by Eq. (2.46); the higher epitaxial-resistivity devices do not because of the heating effects associated with the larger drain voltages.

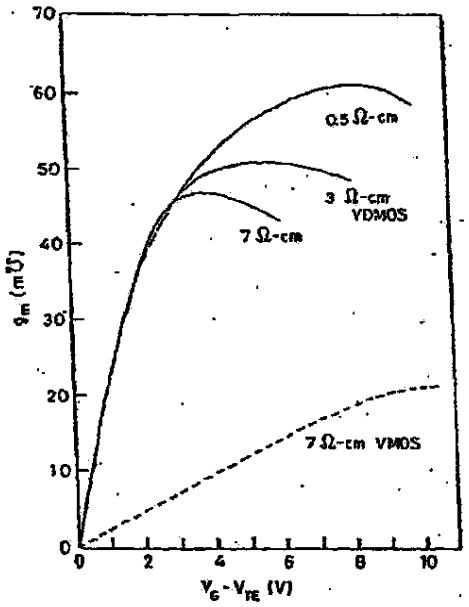
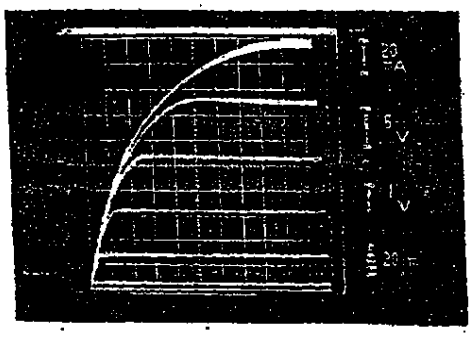


Fig. 2.28. EXPERIMENTAL TRANSCONDUCTANCE MEASUREMENTS ON VDMOS AND VMOS LARGE-GEOMETRY DEVICES.

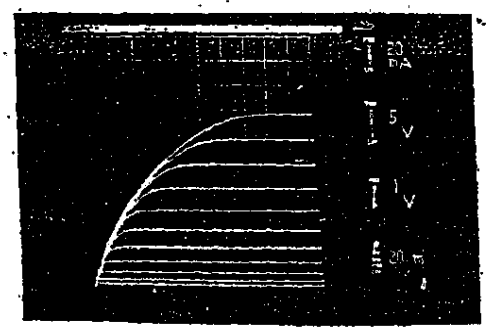
The VMOS has a longer effective channel length (≈1.5 times longer in the simultaneously fabricated devices considered here) and lower field mobility and scattering-limited velocity as the result of its <111>

surface orientation. The VMOS in Fig. 2.29 has much lower transconductance at low gate voltages, therefore, and requires much higher gate voltages to achieve scattering-limited velocity operation. Because the thermal effects are so dominant in the VMOS, it does not obtain a maximum g_m comparable to the VDMOS.

Figure 2.29 compares the I-V characteristics of the 7 Ω-cm large-geometry (W = 3440 μ) VDMOS and VMOS. The difference in transconductance is apparent as is the higher gate voltage in the VMOS required to achieve scattering-limited velocity operation.



a. VDMOS, 7 Ω-cm



b. VMOS, 7 Ω-cm

Fig. 2.29. COMPARISON OF THE EXPERIMENTAL I-V CHARACTERISTICS OF VDMOS AND VMOS DEVICES WITH IDENTICAL CHANNEL WIDTHS.

Similar behavior has been observed in the LDMOS. Figure 2.30 plots the results of measurements obtained from three small-geometry (W = 200 μ) LDMOS devices, using 80 μsec 1 percent duty-cycle pulsed measurements on a standard curve tracer. In the 0.5 Ω-cm material, the measured $g_{m(max)}$ = 4.5 m℧ which is the theoretical value derived in Eq. (2.46) under the velocity-saturation condition. At greater resistivities (such as 3 and 7 Ω-cm), the high channel temperature degrades g_m before the theoretical maximum value is reached; it becomes worse at higher $V_G - V_T$ because a larger V_{DS} is required to maintain the device in saturation and, as a result, more heat is generated in the active channel region.

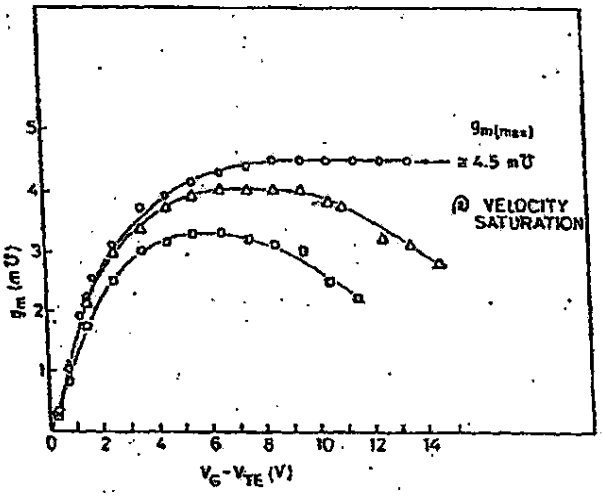


Fig. 2.30. EXPERIMENTAL TRANSCONDUCTANCE MEASUREMENTS FROM THE LDMOS SMALL-GEOMETRY DEVICES.

Using shorter pulses to reduce device heating improves the apparent transconductance as illustrated in Fig. 2.31 where these measurements were obtained from the 3 Ω-cm small-geometry (W = 200 μ) VDMOS. Even with 1 μsec pulses, however, the device does not achieve a transconductance equal to that predicted by Eq. (2.46).

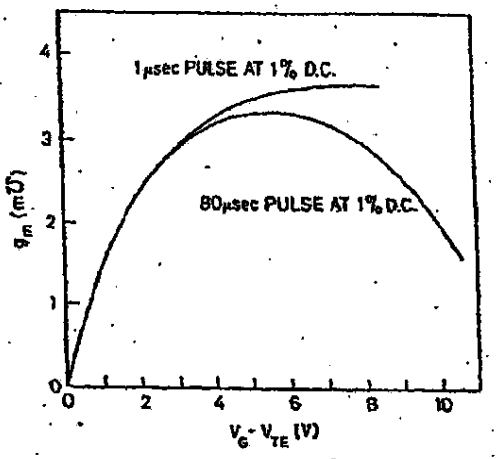


Fig. 2.31. EFFECT OF PULSEWIDTH ON MEASURED TRANSCONDUCTANCE IN THE 3 Ω-cm VDMOS.

It has been postulated [2.33] that heating is localized in the channel region and, because this region is small, its response time is so rapid that small duty cycles do not eliminate the thermal effect but can only prevent the global temperature rise in the substrate. The thermal effect is also assumed to be responsible for the negative resistance in $\partial I_D/\partial V_{DS}$ (Fig. 2.29) of a typical 7 Ω-cm VDMOS; this resistance becomes increasingly more prominent as V_G (and, therefore, I_D) is raised for a given V_{DS}. An alternative explanation [2.22] is the spreading depletion region under the gate when the device is in the saturation region at high V_{DS}.

Figure 2.32 is a schematic of the test circuit† used to measure drain currents with input pulses applied to the gate. The electrometer will produce the current reading which can then be converted to the peak drain current from the equivalent circuit in Fig. 2.33. In steady state, $Q_{charging} = Q_{discharging}$ in the capacitor; therefore,

†Obtained with the assistance of Dr. J. T. Walker, Integrated Circuits Laboratory, Stanford University.

$$I_1 t_{off} = (I_D - I_1)\, t_{on} \qquad RC >> t_{on} + t_{off} \qquad (2.47)$$

which yields

$$I_D = \frac{t_{ON} + t_{off}}{t_{on}} I_1 = \frac{\text{electrometer reading}}{\text{duty cycle}} \qquad (2.48)$$

The voltage across the device V_{DS} is equal to $V_{DD} - I_1 R$. Because $RC >> t_{on} + t_{off}$, the ripples in the capacitor voltage (V_{DS}) are very small. The device was mounted on a TO-3 package similar to the one in Fig. 2.34 [2.34], which provides a good heat sink for high-power devices with gold-germanium alloy die attachments.

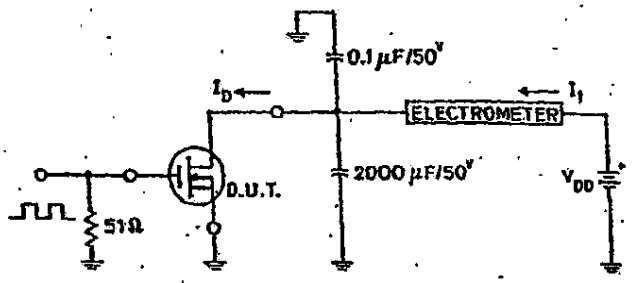


Fig. 2.32. TEST CIRCUIT USED FOR THE MEASUREMENT OF PULSED DRAIN CURRENTS.

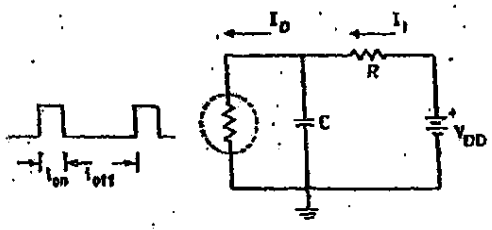


Fig. 2.33. EQUIVALENT CIRCUIT.

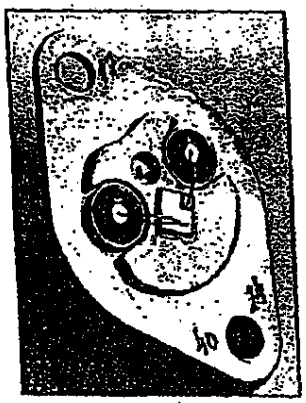


Fig. 2.34. A TO-3 PACKAGED DEVICE.

H. High-Voltage VDMOS I-V Characteristics

The I-V characteristics of power MOSFETs were described briefly in the preceding section. Some of them will now be discussed in greater detail.

Modeling of the LDMOS as an enhancement-mode transistor in series with a bulk resistor has been reported extensively [2.13,2.21,2.35]. For VMOS and low-voltage VDMOS devices, this approach can also apply because bulk resistance is negligibly affected by applied drain bias. In the high-voltage VDMOS, however, the depletion-region widths surrounding the p-wells vary with applied bias because of the voltage drop along the channel of the JFET. As a result, modeling the I-V behavior of these devices becomes extremely complicated. The conventional use of a constant resistor is not applicable because bulk resistance is no longer a constant.

To obtain a cylindrical approximation of the diffused junction requires a numerical method to solve self-consistent equations even with a simplified one-dimensional approach. For example, to complete the I-V relationship for the JFET in Fig. 2.35, the following three equations must be solved numerically [2.36]:

$$\frac{dV}{dx} = \frac{\rho I_D}{hW} \tag{2.49a}$$

$$L - \sqrt{r^2 - x^2} = \frac{h}{2} \tag{2.49b}$$

$$V(x) = \frac{qN_D}{4\varepsilon_{si}} x_j^2 \left[1 - \left(\frac{r}{x_j}\right)^2 + \ln\left(\frac{r}{x_j}\right)^2 \right] \tag{2.49c}$$

where $V(x)$ denotes voltage as a function of vertical position in the JFET channel, and the geometrical dimensions h, L, and r are shown in Fig. 2.35. Equation (2.49c) is derived in Appendix B.

To determine the VDMOS I-V characteristics, these JFET equations also must be solved along with an MOS enhancement-mode device and a bulk resistor whose value depends on the JFET channel thickness. Because this is a very complicated computational process, a different approach was chosen.

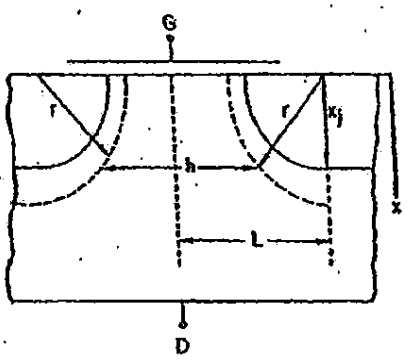


Fig. 2.35. MODEL OF THE VDMOS USED TO OBTAIN THE VOLTAGE ALONG THE JFET CHANNEL.

The required analytical solutions were based on the following assumptions.

- Depletion-mode device resistance is neglected. In high-voltage structures (≥7 Ω-cm epitaxial material), this resistance is less than 5 percent of total resistance when $V_G - V_T > 5$ V.
- The conventional MOS drain-current equation is used, and the velocity-saturation effect is neglected. This effect can be added later to improve the accuracy of the model but at the expense of mathematical complexity. Mobility reduction resulting from strong gate fields, however, is taken into account [2.5].
- After the MOS device pinches off, the drain current remains constant, and the heating effect on mobility caused by power dissipation [2.33] is not considered.
- The abrupt-junction approximation is applied for the p^+-n^- channel drain junction.
- The region between the p-wells is regarded as a junction field-effect transistor, and a rectangular geometry is necessary to obtain one-dimensional analytical solutions.
- The gradual-channel approximation assumes that the equipotential planes are in parallel with the device surface and that current density is uniform and normal to these planes.

- After the JFET pinches off, the current remains constant at the pinchoff value.
- The epitaxial bulk-resistance calculation is based on Eq. (2.26).

The impact of these assumptions when deriving the current-voltage relationship is discussed in this section. The development of the analytical model based on the structure in Fig. 2.36a and the equivalent circuit in Fig. 2.36b proceeds as follows.

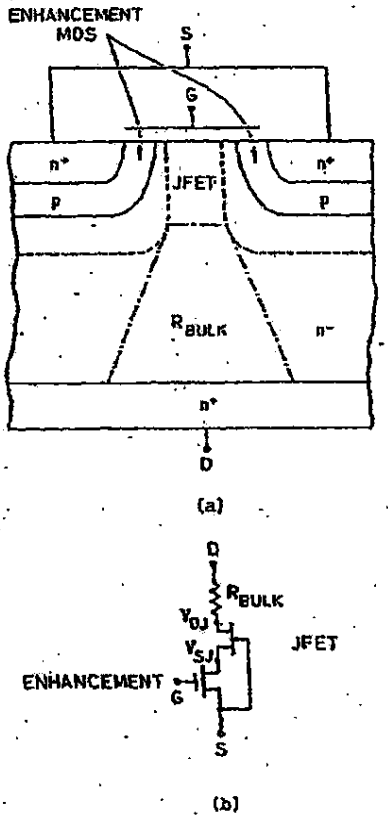


Fig. 2.36. MODEL OF THE VDMOS USED IN THE CALCULATION OF THE I-V CHARACTERISTICS.

The MOS enhancement-mode device is first considered. The standard equation for MOS drain current with zero back-gate bias [2.37] is

$$I_D = \frac{W}{L_{eff}} \mu C_o \left\{ \left(V_{GS} - 2\phi_F - \phi_{ms} + \frac{Q_f}{C_o} \right) V_{DS} - \frac{V_{DS}^2}{2} - \frac{2}{3} \gamma \left[(V_{DS} + 2\phi_F)^{3/2} - (2\phi_F)^{3/2} \right] \right\} \qquad (2.50)$$

Where

W = channel width

L_{eff} = active channel length

V_{DS} = drain-source voltage

V_{GS} = gate-source voltage

ϕ_F = Fermi potential of channel peak doping concentration

γ = bulk-charge factor = $(2\varepsilon_{si} q N_A)^{1/2}/C_o$

ϕ_{ms} = metal-semiconductor work function

N_A = maximum acceptor dopant concentration at the surface of the diffused channel

The drain saturation voltage V_{DSAT} derived in Eq. (2.50) by taking $dI_D/dV_{DS} = 0$ [2.37] is

$$V_{DSAT} = V_{GS} - 2\phi_F - \phi_{ms} + \frac{Q_f}{C_o} + \left(\frac{\gamma}{2}\right)^2 \cdot \left[1 - \sqrt{1 + \frac{4(V_{GS} - \phi_{ms} + Q_f/C_o)}{\gamma^2}} \right] \qquad (2.51)$$

and the saturation drain current I_{DSAT} can be calculated by substituting V_{DSAT} in Eq. (2.51) for V_{DS} in Eq. (2.50).

The current-voltage relation for the JFET as derived below is based on the simplified structure in Fig. 2.37; here, channel thickness 2a

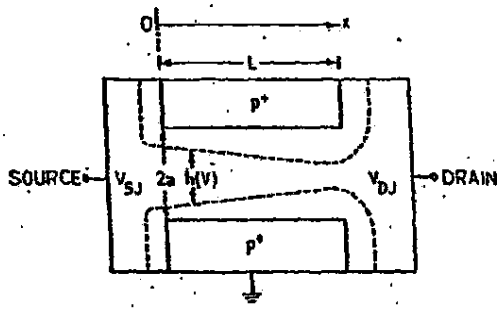


Fig. 2.37. SCHEMATIC OF THE n-CHANNEL JFET USED IN THE MODEL CALCULATION.

is equivalent to L_T in Fig. 2.10. The depletion-layer width of the gate-to-channel junction (using the one-sided abrupt-junction approximation) is

$$X_d = \sqrt{\frac{2\varepsilon_{si}[V(x) + \phi_B]}{qN_D}} \qquad (2.52)$$

where $V(x)$ is voltage as a function of distance along the JFET channel. The gate (which is also the diffused channel of the MOS transistor) is grounded.

The undepleted layer thickness inside the JFET channel is

$$h(V(x)) = 2a - \sqrt{\frac{2\varepsilon_{si}[V(x) + \phi_B]}{qN_D}} = 2\left(1 - \sqrt{\frac{V(x) + \phi_B}{V_p}}\right) \qquad (2.53)$$

where $V_p = qN_D a^2/2\varepsilon_{si}$.

Current density $J(x)$ at point x in the channel is related to the electric field dV/dX by

$$J = \frac{1}{\rho}\frac{dV}{dX} \qquad (2.54a)$$

and current density is

$$J = \frac{I_D}{Wh(V)} \qquad (2.54b)$$

When combined, these expressions yield

$$h(v)\, dV = \frac{\rho I_D}{W}\, dx. \qquad (2.55)$$

The current can be determined by integrating Eq. (2.55) from $X = 0$ to $X = L$ on the right and from $V = V_{SJ}$ (source end) to $V = V_{DJ}$ (drain end) on the left in Fig. 2.37:

$$\int_{V_{SJ}}^{V_{DJ}} \left(1 - \sqrt{\frac{V + \phi_B}{V_p}}\right) dV = \int_0^L \frac{\rho I_D}{2aW}\, dx \qquad (2.56)$$

After integrating the JFET drain current,

$$I_{DJ} = \frac{W2a}{L\rho}\left\{(V_{DJ} - V_{SJ}) - \frac{2}{3}\left(\frac{1}{V_p}\right)^{1/2} \cdot \left[(V_{DJ} + \phi_B)^{3/2} - (V_{SJ} + \phi_B)^{3/2}\right]\right\} \qquad (2.57)$$

where, in the JFET,

L = channel length

V_{DJ} = drain voltage

V_{SJ} = source voltage and also the drain voltage of the MOS transistor

This expression is similar to the standard JFET current equation [2.38] but differs in that the gate voltage to the JFET is not generated by an external bias; instead, it originates in the drain voltage of the MOS transistor (source voltage V_{SJ} of the JFET), and voltage drops along the JFET channel while its gate is grounded. The undepleted channel thickness at the drain end is obtained from Eq. (2.53) as

$$h(V_{DJ}) = 2\left(1 - \sqrt{\frac{V_{DJ} + \phi_B}{V_p}}\right) \qquad (2.58)$$

and the JFET pinches off when $h(V_{DJ}) = 0$. As a result, its pinchoff voltage becomes

$$V_{DJ(pinchoff)} = V_p - \phi_B \qquad (2.59)$$

The pinchoff current can be determined by substituting Eq. (2.59) for V_{DJ} in Eq. (2.57).

For a given device geometry, the gate voltage of the VDMOS at which the MOS and JFET pinch off simultaneously ("crossover" point) can be calculated by solving

$$I_{DSAT}(MOS) = I_{D(pinchoff)}(JFET) \qquad (2.60)$$

Referring to Fig. 2.36, the V_{SJ} of the JFET here is the V_{DSAT} of the MOS transistor.

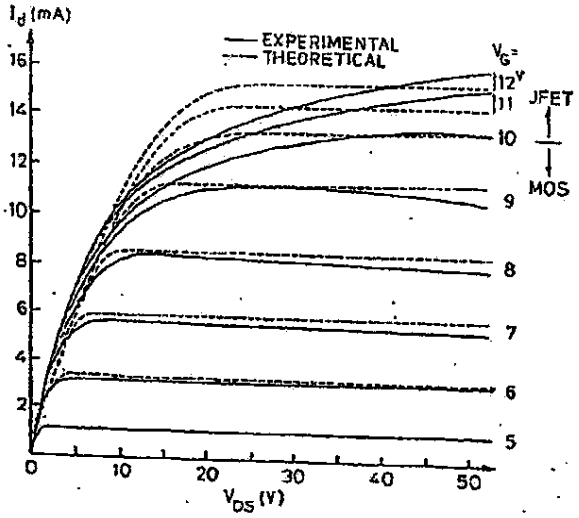
When the applied gate voltage to the VDMOS is less than the crossover voltage [solution in Eq. (2.60)], the current flowing in the JFET transistor is not sufficient to result in pinchoff. Because the MOS device will saturate before JFET pinches off, the VDMOS current is controlled by the MOS component in Fig. 2.36 and the VDMOS I-V characteristics are similar to an MOS enhancement-mode transistor. When the applied gate voltage is greater than the crossover voltage, the higher current flow will cause JFET to pinch off before the MOS transistor saturates. Under this condition, the MOS component is in the linear region and its behavior is similar to a resistor. The VDMOS current-voltage characteristics are dictated by the JFET component in Fig. 2.36.

The key geometrical parameter in determining the crossover gate voltage is the spacing between the p-wells. For narrow spacings, the JFET will dominate VDMOS behavior as $V_G - V_T$ increases; for wider spacings, the enhancement-mode MOS could dominate over a wide range of $V_G - V_T$ and is only limited by heat dissipation. The JFET can be considered, therefore, as a resistor.

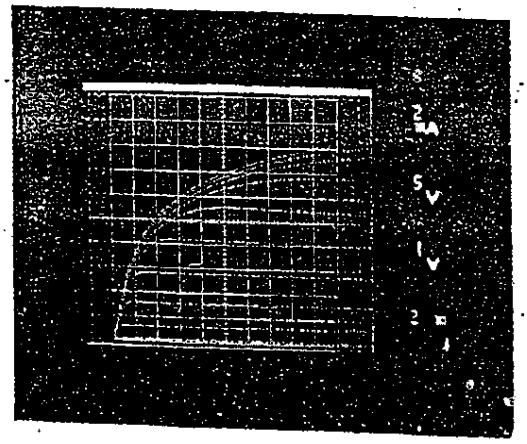
To complete the VDMOS I-V relationship, the external drain voltage was obtained by adding the voltage drop across the bulk resistor to V_{DJ} in Fig. 2.36. The current in this resistor is the same as in the JFET, and the bulk-resistance value was calculated from Eq. (2.26).

The I-V relationship described above was applied to the 3 Ω-cm small-geometry (W = 200 μ) VDMOS whose epitaxial-layer thickness and L_{eff} are listed in Table 2.1. The p-well junction depth was 4.5 μ and its gate-oxide thickness was 950 Å. The mask spacing between the p-wells was 15 μ. Inversion-layer mobility data were obtained from Ref. 2.5. The crossover gate voltage was first determined to be approximately 10 V, below which device current was calculated via Eq. (2.50) for a given MOS drain voltage V_{DS}. This V_{DS}, in turn, was substituted for V_{SJ} in Eq. (2.57), and the same I_D (MOS) was used for I_{DJ} (JFET) to solve for V_{DJ}. After V_{DJ} is known, the bulk resistance and voltage drop can be calculated as above. Beyond MOS saturation, the current is assumed constant at the saturation value. If the gate voltage is greater than the crossover voltage, the MOS transistor will be in the linear region. The JFET drain voltage V_{DJ} can be calculated, starting with the MOS-transistor I-V equation. If this voltage is less than the pinch-off value {Eq. (2.59)}, the VDMOS remains in the linear region; if it is equal to or greater, the current obtained in Eq. (2.59) will stay constant. The external drain bias of the VDMOS is obtained in the same way as those whose gate voltage is less than the crossover value.

The calculated I-V characteristics thereby obtained are plotted in Fig. 2.38a and, for comparison, the curve tracer of 80 μsec pulsed measurements is shown in Fig. 2.38b. The curves correspond to a spread in V_G from 5 to 13 V in 1 V steps, and the threshold voltage was approximately 3.5 V. The agreement between theory and experiment is reasonably good. At gate voltages of less than 10 V, VDMOS displays a typical DMOS I-V characteristic where the drain current remains essentially constant after saturation; at gate voltages of greater than 10 V, the JFET characteristics dominate. Transconductance drops sharply, and the transition from the linear to saturation region is more gentle than when the MOS transistor controls VDMOS behavior.



a. Comparison to theory



b. Measured I-V characteristics

Fig. 2.38. HIGH-VOLTAGE VDMOS (3 Ω-cm, L_T = 15 μ) I-V CHARACTERISTICS.

1. Theoretical Limitations

The observed discrepancies between experiment and theory in the MOS and JFET regions are the result of the following conditions.

- In the MOS mode, velocity saturation was not considered, which resulted in a higher calculated drain current.
- In the JFET mode, the calculated drain current was slightly higher than the experimental data because the depletion-mode device resistance was omitted in the derivations. The voltage drop across this resistance caused negative feedback on the JFET gate-to-source voltage and lowered the current; however, this behavior was compensated for by rectangular geometry which should result in a lower drain current and smaller pinchoff voltage compared to the diffused junction which is more close to cylindrical geometry.
- The large saturation conductance observed in the experimental curve is the result of JFET channel-length modulation which was not accounted for in the calculations. Figure 2.39 is a cross section of the device profile beyond saturation where the pinchoff region moves toward the source as the drain voltage is increased and the effective channel length is shortened. This results in an upward tilt of the I-V characteristics beyond saturation which is particularly most apparent in devices with small channel lengths [2.39].

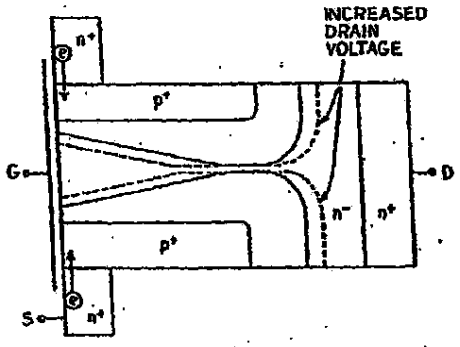


Fig. 2.39. CROSS SECTION OF CHANNEL-LENGTH MODULATION IN THE JFET PORTION OF THE VDMOS.

- Heating also causes the JFET saturation current to drop as the drain voltage is increased (evident at V_G = 10 V in Fig. 2.36). The major effects of temperature on a JFET are to alter the average channel conductivity [2.40] and to raise the reverse gate bias on the JFET as a result of higher MOS channel resistance.

- Such second-order conditions as variation of mobility with field in the JFET [2.41] should produce a lower saturation drain current compared to the first-order approximation, but this was not considered in the calculations.

The effect of spacing L_T between the p-wells on the I-V behavior is illustrated in Fig. 2.40. The 80 μsec pulsed measurements were obtained from an L_T = 40 μ (mask dimension) small-geometry (W = 200 μ) VDMOS. Because this device was fabricated on the same wafer as the device (L_T = 15 μ) in Fig. 2.38, its threshold voltage is also approximately 3.5 V. Over the same spread of V_{GS} = 5 to 13 V, there is no sign of JFET dominance; the enhancement-mode MOS controls the I-V characteristics over the entire V_{GS} range. It should be noted that the vertical scale is 5 mA/div.

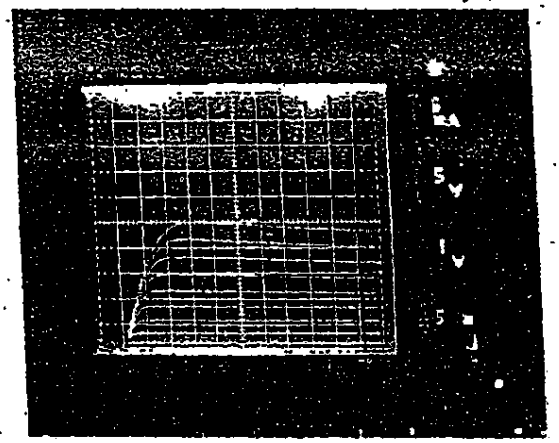


Fig. 2.40. VDMOS (3 Ω-cm, L_T = 40 μ) I-V CHARACTERISTICS.

2. Limitations of Device Transconductance

When the high-voltage VDMOS operates in the MOS mode (with the correct spacings between the p-wells), transconductance can be described by Eqs. (2.45) and (2.46). When L_T is narrower, however, there are two distinct MOS and JFET modes of operation. The behavior of device transconductance is described in Fig. 2.41. At low $V_{GS} - V_T$ where MOS dominates, g_m is proportional to $V_{GS} - V_T$; at higher $V_{GS} - V_T$, the effects of mobility reduction and velocity saturation appear and g_m deviates from its initial slope. The maximum value of g_m is limited by velocity saturation [Eq. (2.38)] and temperature. As $V_{GS} - V_T$ reaches the crossover point where JFET starts to dominate, g_m begins to fall. At very small L_T, the JFET can take control even before the MOS-transistor g_m reaches the limited value set by velocity saturation.

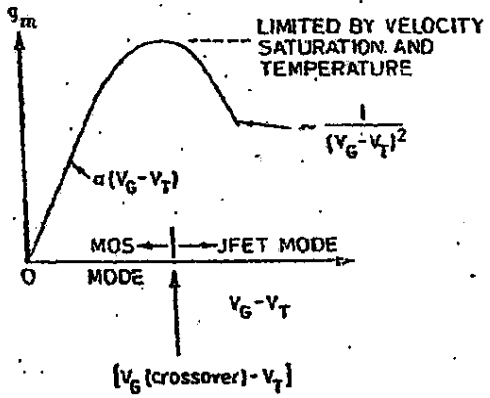


Fig. 2.41. NARROW-SPACING VDMOS TRANSCONDUCTANCE.

The functional dependence of VDMOS transconductance in the JFET mode on $V_{GS} - V_T$ can be determined by

substituting

$$V_{DS} = \frac{I_D}{(W/L_{eff})\,\mu C_o (V_{GS} - V_T)}$$

for V_{SJ} in Eq. (2.57) because MOS drain voltage V_{DS} is also JFET source voltage V_{SJ} (see Fig. 2.36)

- neglecting the last term in Eq. (2.57) because V_{SJ} is relatively small compared to V_{DJ}
- substituting $V_p - \phi_B$ for V_{DJ} because JFET is in the pinchoff region

The drain current now becomes

$$I_{D(pinchoff)} = I_o\left[V_p - \phi_B - \frac{I_D}{\beta(V_{GS} - V_T)} - \frac{2}{3}V_p\right] \qquad (2.61)$$

where

$$I_o = \frac{W2a}{L\rho}$$

$$\beta = \frac{W}{L}\mu C_o$$

Rearranging Eq. (2.61), I_D becomes

$$I_D = \frac{I_o[(1/3)\ V_p - \phi_B]}{1 + [I_o/\beta(V_{GS} - V_T)]} \qquad (2.62)$$

The VDMOS transconductance g_m is defined as

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \qquad (2.63)$$

Differentiating I_D in Eq. (2.62) with respect to V_G yields

$$g_m = \frac{I_D I_o}{\beta(V_{GS} - V_T)^2\left\{1 + [I_o/\beta(V_{GS} - V_T)]\right\}}$$

$$= \frac{I_o^2(1/3\ V_p - \phi_B)/\left\{1 + [I_o/\beta(V_{GS} - V_T)]\right\}}{\beta(V_{GS} - V_T)^2\left\{1 + [I_o/\beta(V_{GS} - V_T)]\right\}} \qquad (2.64)$$

If $I_o/\beta(V_{GS} - V_T) \ll 1$ (that is, $2a/z \ll \mu C_o(V_{GS} - V_T)$) which is typical in the JFET mode, Eq. (2.64) then becomes

$$g_m \approx \frac{I_o^2[(1/3)V_p - \phi_B]}{\beta(V_{GS} - V_T)^2} \qquad (2.65)$$

Because the JFET is deep in the pinchoff region, the change in VDMOS gate voltage V_G causes only a small variation in the MOS-resistance value (the MOS component in Fig. 2.36 is in the linear region). This, in turn, generates little change in the JFET gate bias and, as a result, g_m is severely restricted. An upper limit could be set, therefore, on the usefulness of the VDMOS as a linear amplifier. The other extreme is $I_o/\beta(V_{GS} - V_T) \gg 1$, which is very unlikely in JFET-mode operations; however, if this condition holds, Eq. (2.64) becomes $g_m \approx \beta(V_p - \phi_B)$.

The experimental data obtained from the curve tracers in Figs. 2.38b and 2.40 are plotted in Fig. 2.42. The behavior of both devices is similar to the theoretical predictions. Because of the narrower spacing between the p-wells, g_m in the 15 μ device begins to fall at

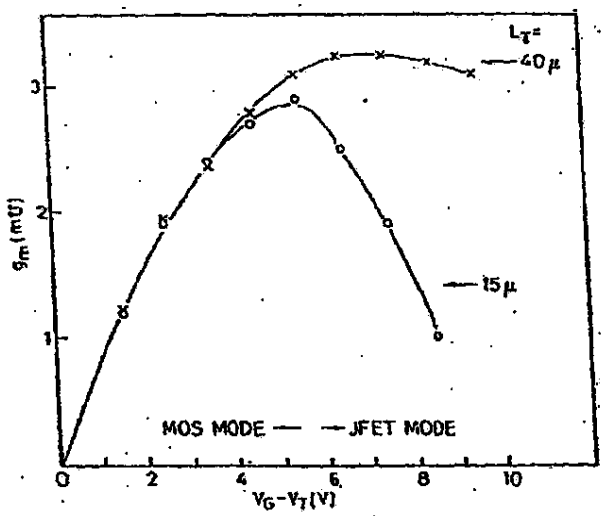


Fig. 2.42. EXPERIMENTAL TRANSCONDUCTANCE VS GATE DRIVE.

$V_{GS} - V_T > 5.5$ V--a typical JFET mode of operation. In the 40 μ device, there is no indication of the JFET mode within the $V_{GS} - V_T$ range; instead, it exhibits a typical DMOS transconductance characteristic. The small decrease of g_m at high $V_{GS} - V_T$ is probably the result of the heating effect.

As discussed in Section F, on-resistance decreases as device geometry becomes small. A critical parameter in scaling is the aspect ratio of p-well spacing to p-well junction depth. Incorrect scaling could limit transconductance severely at high gate voltages which, in turn, would restrict the current handling capability. It may be necessary to optimize the device for specific applications.

An interesting variation of the VDMOS is a structure called buried-load logic [2.42] used in logic applications (Fig. 2.43). Except for an N^+ diffusion between the p-wells and an added contact to this diffusion, it is identical to the model described in Fig. 2.36. The I-V analysis developed in this section is also applicable to this structure.

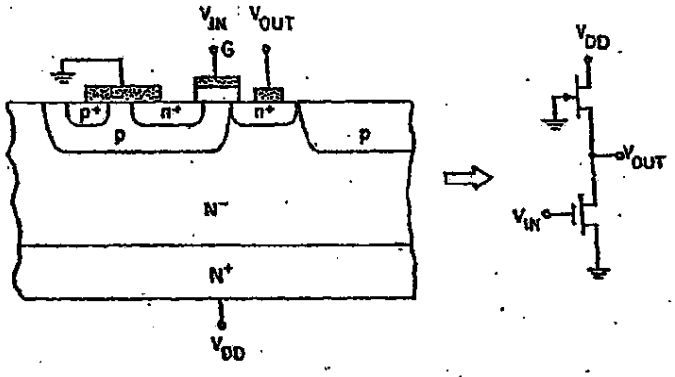
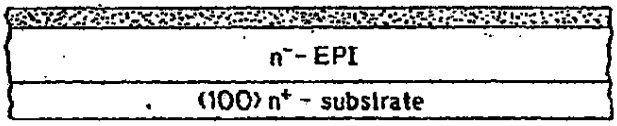


Fig. 2.43. BURIED-LOAD LOGIC--A VARIATION OF THE VDMOS STRUCTURE.

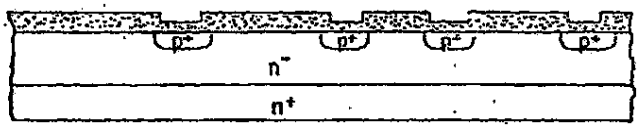
I. Device Fabrication

A planar DMOS process [2.3] was modified to fabricate the LDMOS, VDMOS, and VMOS simultaneously on the same chip. Highly antimony-doped N^+ <100>-oriented substrates were used, and N^- layers with resistivities

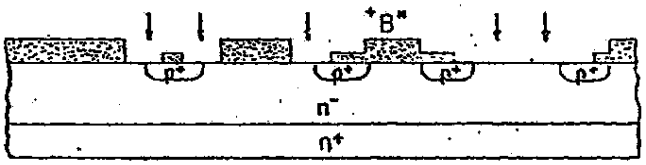
ranging from 0.5 to 8.5 Ω-cm were grown via conventional two-step epitaxial techniques. Following epitaxial growth, the fabrication process included P^+ channel contact diffusion, P^- channel implant and drive-in, N^+ source-and-drain diffusion (for the LDMOS), V-groove etching (for the VMOS), gate oxidation, contact-hole opening, metal deposition and definition, and passivation. The processing steps and schedule are illustrated and outlined in Fig. 2.44 and Table 2.3.



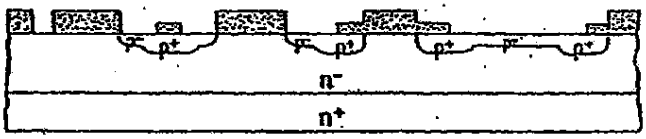
a. Initial oxidation



b. After P^+ contact drive-in

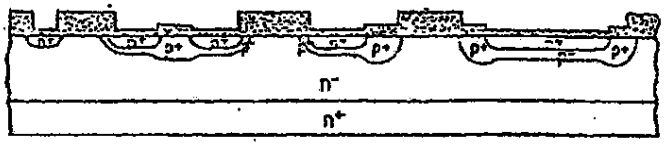


c. Channel implant and drive-in

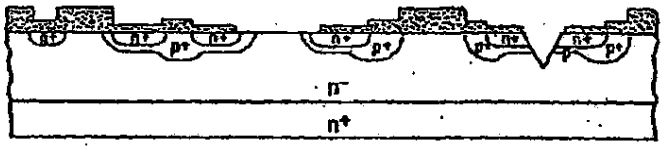


d. Drain opening

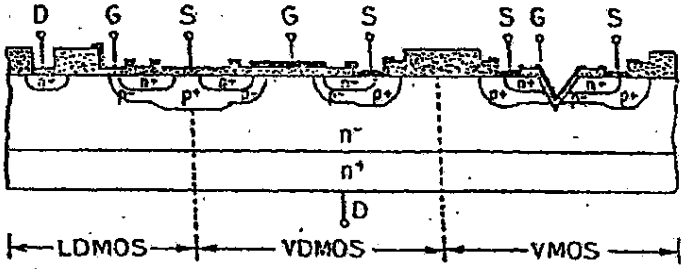
Fig. 2.44. PROCESSING STEPS FOR THE LDMOS, VMOS, AND VDMOS.



e. After source and diffusion and oxidation



f. After V-groove and gate openings



g. After metal photolithography

Fig. 2.[illegible]. CONTINUED.

Table 2.3

LDMOS, VDMOS, AND VMOS PROCESS SCHEDULE

Step	Process
1. Starting wafer	<100> N-type, 0.005 to 0.015 Ω-cm, antimony doped
2. Epitaxy	N-type, arsenic doped, thickness and resistivity according to Eq. (2.3)
3. Initial oxidation	1000°C, SDO_2-$26SWO_2$-1000_2, t_{ox} = 1 μm

Table 2.3

CONTINUED

4. P^+ contact photolithography	etch time in 6:1 buffered HF (BOE) = 10 min
5. P^+ contact predeposition	diborane: 30 min @ 1000°C, R_S = 24 - 30 Ω/□
6. P^+ contact drive-in	1050°C, $5DO_2$-$50WO_2$-$5DO_2$, t_{ox} = 4900 - 5300 Å, R_S = 88 - 115 Ω/□
7. Channel (p-well) photolithography	etch time in BOE = 12 min
8. Channel implant	$^+B^{11}$: 3×10^{14} - 5×10^{14} @ 50 keV
9. Channel drive-in	1200°C, $2DO_2$-$90N_2$, t_{ox} < 500 Å, R_S = 103 - 170 Ω/□
10. Drain photolithography	etch time in BOE = 12 min
11. Source and drain n^+ predeposition	$POCl_3$ @ 4°C, 30 min @ 1000°C
12. n^+ drive-in	1050°C, $10DO_2$-$40WO_2$-$10DO_2$, t_{ox} = 4450 Å, R_S = 10.2 - 15 Ω/□
13. V-groove photolithography	etch in ethylene diamine @ 118°C
14. Gate photolithography	etch time in BOE = 13 min
15. Gate oxidation	950°C, $5DO_2$-$20WO_2$-$10DO_2$-$15N_2$
16. Contact photolithography	etch time in BOE = 7 min
17. Metal deposition	aluminum E-beam evaporation both front and back surfaces, 1.5 μ
18. Metal photolithography	etch in phosphoric/nitric-acid mixture @ 45°C
19. Silox deposition	5000 Å phosphorus 1 - 3% doped oxide, 2000 Å undoped @ 450°C
20. Passivation photolithography	etched in Vapox etch

The important process considerations are as follows.

- An antimony-doped rather than arsenic-doped N^+ substrate was chosen because of the autodoping associated with high arsenic-vapor pressures during epitaxial-layer growth.
- A 1000°C initial oxidation temperature was used instead of the standard 1200°C to minimize outdiffusion of the N^+ substrate. Both the antimony diffusion coefficient D and silicon-dioxide growth-rate constant B_1 are functions of temperature. The activation energy for D is 3.98 [2.43] and is ≈0.78 [2.44] for B_1; therefore, D decreases faster than B_1 when the oxidation temperature is lowered from 1200° to 1000°C. Although the same amount of oxide can be grown at a lower temperature and a longer oxidation cycle as at a higher temperature and shorter cycle, the ratio of the DT product (T = temperature) is significant—DT (90 min @ 1200°C)/DT (280 min @ 1000°C) = 49. The square root of DT is a measure of the extent of outdiffusion. The impact of this change in temperatures on the antimony outdiffusion profile as observed in the SUPREM simulation [2.43] is plotted in Fig. 2.45.
- Ion implantation was used in the channel predeposition to better control the MOS threshold voltages.
- Gate-oxide thickness was 950 Å in the LDMOS and VDMOS and 1350 Å in the VMOS because of the different growth rates on the <100> and <111> surfaces [2.8].

The threshold voltages of DMOS transistors for different epitaxial-layer resistivities and for a 3×10^{14} ions/cm^2 implant dose at 50 keV are

EPI resistivity (Ω-cm):	0.5	1.12	3.0
V_T (V):	1.94	2.7	2.8

It is apparent that epitaxial doping at 0.5 Ω-cm has lowered the threshold appreciably because the peak channel doping concentration is approximately 5 to 6×10^{16} cm^{-3} after gate oxidation. The epitaxial-layer doping concentration at 0.5 Ω-cm is $\approx 1.2 \times 10^{16}$ cm^{-3}. In addition, the phosphorus pile-up near the silicon surface will compensate for the net acceptor surface concentration. The threshold voltages of the 0.5 Ω-cm epitaxial material for different implant doses are

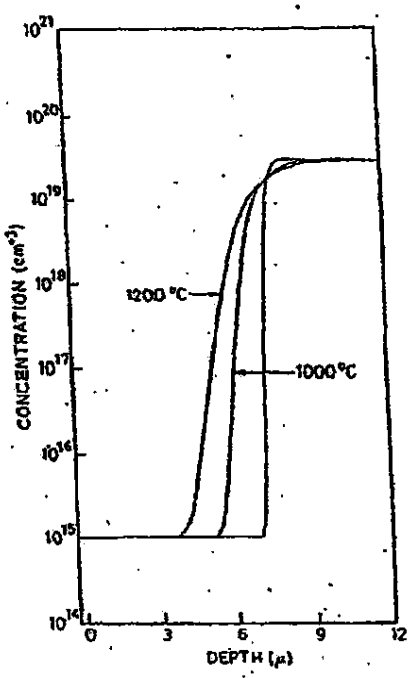


Fig. 2.45. EFFECT OF INITIAL-OXIDATION TEMPERATURES ON N^+ SUBSTRATE OUTDIFFUSION.

Implant dose (ions/cm^2):	3×10^{14}	4×10^{14}	5×10^{14}
V_T (V):	1.94	2.50	3.11

Because sequential diffusions of P and N^+ impurities are used to form short channel lengths, the vertical impurity profile is similar to a double-diffused bipolar transistor. The peak current gain β obtained by the 0.5 Ω-cm epitaxial material and a 3×10^{14} ions/cm^2 implant dose is roughly 20. Here, $BV_{CBO} = 59$ V and $BV_{CEO} = 36$ V. If $BV_{CEO} \approx BV_{CBO}/\beta^{1/n}$ (2.45) is used to relate BV_{CEO} to BV_{CBO}, the n factor will be 6.

As a result of double diffusion, the P^- channel region under the n^+ source becomes a pinched resistor. The pinched resistance of a 3×10^{14} ions/cm^2 implant dose with no source/drain bias is

R_S = 3.8 kΩ/□ (0.5 Ω-cm EPI)

R_S = 3.0 kΩ/□ (1 ~ 3 Ω-cm EPI)

J. Summary

Several planar and nonplanar DMOS structures have been investigated for power-MOS transistor applications, with emphasis on the LDMOS, VDMOS, and VMOS devices fabricated side by side on the same wafer. Comparisons of these devices indicated that the differences in on-resistance and transconductance are to be expected because of the variations in the physical structures and in such basic material parameters as mobility and scattering-limited electron velocity. Models relating on-resistance to the structures were developed and were found to be in good agreement with experimental data. These models should enable direct comparisons of the devices for a given application when technological and design-rule parameters are specified.

In low-voltage applications, the VDMOS and LDMOS have lower on-resistance per unit width than does the VMOS; at high voltages, all three are approximately equal. Technological considerations and layout efficiency may be more important, however, in determining which device produces lower overall on-resistance for a given chip area. Transconductance is generally higher in the LDMOS and VDMOS than in the VMOS for equal channel lengths because of higher electron mobilities and scattering-limited velocities on the <100> crystal plane. Thermal effects make it difficult to achieve room-temperature scattering velocity-limited transconductance, particularly at higher voltages; in high-power applications, they may also determine on-resistance and current capability. This negative temperature coefficient, however, makes these power MOSFETs relatively immune to thermal runaway and secondary breakdown problems.

Models relating the VDMOS I-V characteristics to the physical structures were also developed and agree well with experimental data. The parasitic JFET could set another limitation on transconductance in addition to the thermal effects.

Chapter III

LIMITATIONS OF POWER MOSFETs

Because the power MOSFET has the surface structure of an MOS and the bulk structure of a bipolar device, the breakdown mechanisms in both types exist in the power MOS transistor. It is important, therefore, to study the possible breakdown limitations that can occur to determine which are most important in a given application.

This chapter first discusses the gate-dielectric, punchthrough, and junction-edge avalanche breakdown limitations. The static breakdown limit (BV_{CEO}) in the parasitic bipolar device (intrinsic to DMOS transistors) is then examined as is the limit imposed on the power MOSFET switching behavior (dv/dt). The key parameters that affect device performance are identified. The Johnson limit [3.1] is the ultimate device figure of merit and is described for both MOS and bipolar transistors.

A. Dielectric Breakdown Limit

1. Gate-to-Source Breakdown

Gate overvoltage damage is a common problem in all MOS devices because the thin insulating gate oxide is subject to overvoltage failures. The intrinsic oxide breakdown field for 1000 Å oxide is $\approx 9 \times 10^6$ V/cm and, below 400 Å, this maximum field is $\approx 1.1 \times 10^7$ V/cm and is slightly temperature-dependent [3.2]. The practical oxide breakdown field is much lower, however, and is more processing-dependent as the result of substrate defects, aluminum migration, or sodium contamination [3.3]. The normal operating field should be much less than the intrinsic value; a reasonable limit is ≈30 to 40 V/1000 Å.

2. Gate-to-Drain Breakdown

Gate-to-drain breakdown voltage in the LDMOS can be greatly reduced when the gate is extended toward the n^+ drain contact region [3.4]; however, the gate overlaps only the n^- drift region of the drain in the vertical power MOSFETs. This n^- region consumes a large portion of applied bias, which prevents high electric-field exposure of the gate

oxide, as will be explained in this section. Because of the two-dimensional nature of the VMOS, a simple one-dimensional analytical solution for the gate-oxide field at a specific drain voltage is not possible; a two-dimensional solution of Poisson's equation is a more successful approach [3.5]. In the VDMOS and TVMOS, the maximum electric field appears in the center of the gate region as illustrated in Fig. 3.1. The electric fields in the oxide and on the silicon surface can be determined from a one-dimensional solution of Poisson's equation when the region under the gate is depleted as normally occurs at high applied V_{DS}.

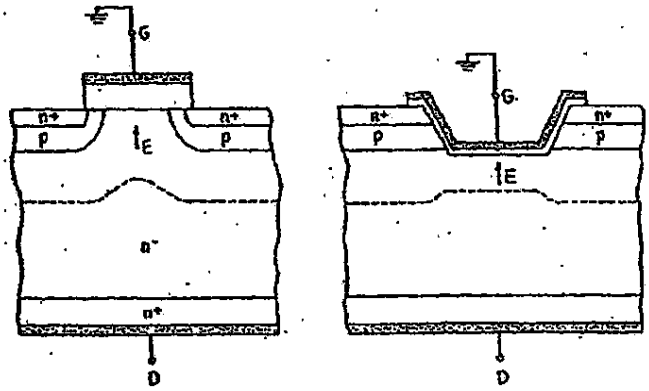


Fig. 3.1. ONE-DIMENSIONAL APPROXIMATION OF THE ELECTRIC FIELD UNDER THE GATE REGION IN THE VDMOS AND TVMOS.

When the gate is grounded and a positive voltage is applied to the drain, mobile carriers will be depleted from the silicon surface. This is equivalent to when the drain is at the reference point (ground) and negative bias is applied to the gate. An energy-band diagram in a plane perpendicular to the channel (Fig. 3.2) facilitates the computation of the necessary voltages and electric fields.

The voltage across the oxide [3.6] is

$$V_{ox} = -\psi_s + V_G + \phi_{ms} \qquad (3.1)$$

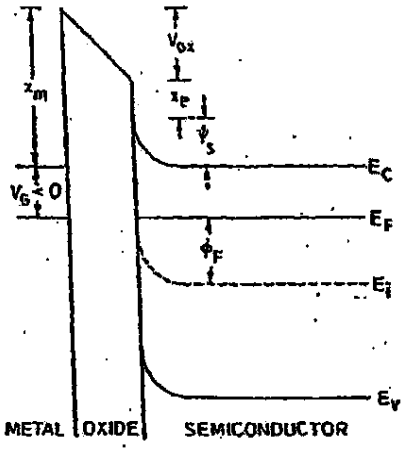


Fig. 3.2. MOS ENERGY-BAND DIAGRAM FOR THE n-TYPE SEMICONDUCTOR UNDER THE DEPLETION CONDITION ($V_G < 0$). The substrate is grounded.

where

ψ_s = surface potential

V_G = gate-to-ground voltage

ϕ_{ms} = metal semiconductor work-function difference

$= \chi_m - (\chi_e + E_g/2q - \phi_F)$

Based on the Gauss law, the field in the oxide E_{ox} is

$$\varepsilon_{ox} E_{ox} = \varepsilon_{si} E_s + Q_f \qquad (3.2)$$

where E_s is the electric field at the semiconductor surface and Q_f is the equivalent SiO_2/Si interface charge density. The total charge per unit area in the semiconductor Q_s is related to E_s by

$$Q_s = \varepsilon_{si} E_s \qquad (3.3)$$

which, when rearranged, yields

$$V_{ox} = t_{ox}E_{ox} = \frac{t_{ox}}{\epsilon_{ox}}(Q_s + Q_f) = \frac{Q_s + Q_f}{C_o} \qquad (3.4)$$

By combining Eqs. (3.1) and (3.4),

$$\psi_s + \frac{Q_s}{C_o} = V_G + \phi_{ms} - \frac{Q_f}{C_o} \qquad (3.5)$$

The semiconductor Q_s is related to substrate doping concentration N_D by

$$Q_s = qN_DX_D \qquad (3.6)$$

where X_D is the depletion width. By definition, flatband voltage V_{FB} is

$$V_{FB} = \phi_{ms} - \frac{Q_f}{C_o} \qquad (3.7)$$

Substituting these last two expressions into Eq. (3.5) yields

$$\psi_s + \frac{qN_DX_D}{C_o} = V_G + V_{FB} \qquad (3.8)$$

Under the abrupt-depletion approximation, the relationship between ψ_s and N_D is identical to that of the one-sided abrupt p-n junction [3.7]. After solving the one-dimensional Poisson equation, the surface potential becomes

$$\psi_s = \frac{qN_D}{2\epsilon_{si}}X_D^2 \qquad (3.9)$$

and the electric field at the semiconductor surface becomes

$$E_s = \left(\frac{2qN_D\psi_s}{\varepsilon_{si}}\right)^{1/2} \qquad (3.10)$$

After eliminating X_D from Eqs. (3.8) and (3.9), ψ_s is related to applied bias V_G (3.8) by

$$\psi_s = \left(\sqrt{\alpha^2 + V_G - V_{FB}} - \alpha\right)^2 \qquad (3.11)$$

where

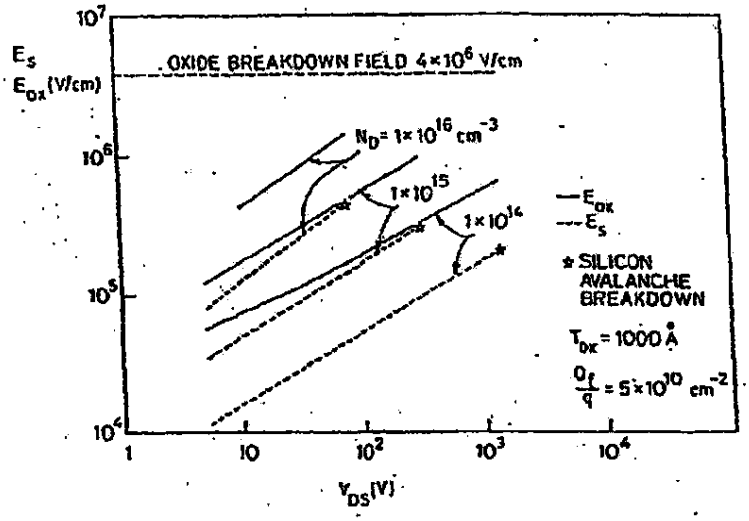
$$\alpha = \frac{\sqrt{qN_D\varepsilon_{si}/2}}{C_o}$$

From Eq. (3.2), the electric field in the oxide is

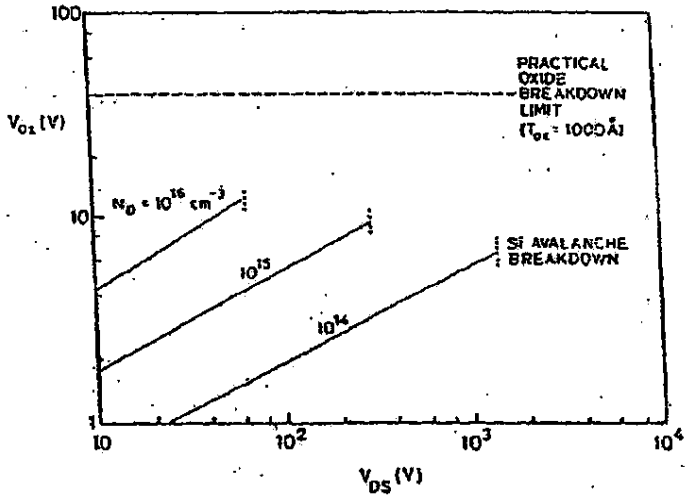
$$E_{ox} = \frac{\varepsilon_{si}E_s + Q_f}{\varepsilon_{ox}} \qquad (3.12)$$

Both E_s and E_{ox} can now be calculated after the applied bias has been specified.

For $t_{ox} = 1000$ Å and $Q_f/q = 5 \times 10^{10}$ cm^{-2} (typical in a <100> MOS), E_{ox} and E_s are plotted in Fig. 3.3a as a function of V_{DS} at several values of N_D, and the variation of V_{ox} with V_{DS} is plotted in Fig. 3.3b. As expected, the voltage drop across the oxide is only a small fraction of the total applied potential at low substrate dopings. Oxide voltage increases as N_D becomes larger; however, E_s will always reach the critical value of avalanche breakdown (Fig. 3.3a) before E_{ox} arrives at its breakdown field. This occurs even in oxides as thin as 100 Å. Because there is a consistent voltage drop across the oxide, the voltage across the depletion region under the gate is less than that across the channel-drain junction. The gate oxide overlapping the drain region in power MOSFETs will never rupture because breakdown will first occur in the semiconductor and is most likely limited by junction curvature and surface effects. Based on Eqs. (3.10) and (3.12), oxide breakdown will begin when N_D is increased to $\approx 4 \times 10^{18}$ cm^{-3} at $t_{ox} \approx 1050$ Å.



a. Oxide and semiconductor surface fields



b. Oxide voltage

Fig. 3.3. OXIDE FIELD AND VOLTAGE VS DRAIN-SOURCE VOLTAGE. The gate is grounded.

The above one-dimensional analysis is considered to be the worst case for possible gate-oxide breakdown. As the p-well spacing is reduced, the overlap of the junction depletion layers caused by the two-dimensional effects will further minimize the gate-oxide field.

B. Punchthrough Breakdown Limit

Channel punchthrough can occur when the depletion layer of the channel-drain (CD) junction reaches the depletion region of the source-channel junction. Although the depletion layer of the CD junction spreads principally into the lightly doped epitaxial region, channel punchthrough could occur as the result of insufficient impurity charge in the channel under strong reverse bias.

The lateral DMOS punchthrough voltages have been calculated in detail [3.9]. The VDMOS and VMOS have the same limitation, but their vertical structures differ from the LDMOS. This section analyzes the punchthrough voltage in vertical FETs with N^+ substrates and under the reachthrough condition.

The impurity profile in the channel strongly influences the final results of the voltage-limitation analysis. Ideally, a gaussian profile more closely resembles the actual profile; however, two integrations are required to obtain the potential across the channel-drain junction. After the first integration, the gaussian function becomes the error function and the second integration becomes nonanalytical. As a result, three approximations (exponential, linear, and step functions) are compared with respect to their differences on the limit.

1. Exponential Approximation of the Channel Profile

Punchthrough voltage can be calculated approximately by assuming an exponential distribution of impurity in the channel diffusion and a uniform distribution in the substrate (Fig. 3.4). Because a portion of the depletion charges on the N^- side is shared by the gate field (Fig. 3.5), the width of the depletion layer near the surface on the channel side will be smaller than that in the bulk. When the N^- depletion layers of the two p-wells are merged, channel depletion becomes less under the gate. Punchthrough most likely occurs, therefore, away from the surface.

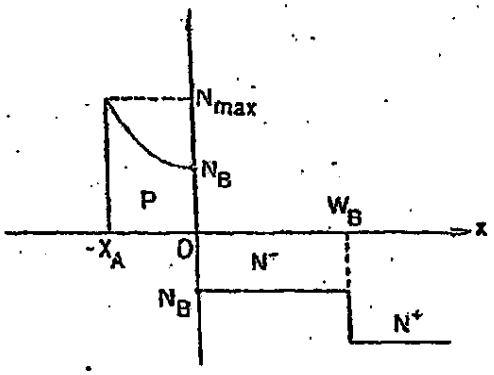


Fig. 3.4. IDEALIZED IMPURITY PROFILE USED IN THE CALCULATION.

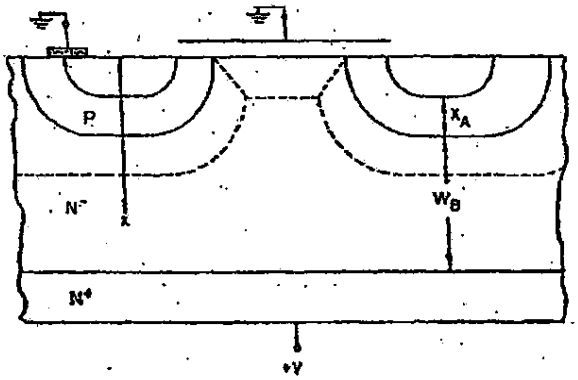


Fig. 3.5. CHARGE SHARING UNDER GATE PUNCHTHROUGH. Punchthrough is not likely to occur near the surface.

Based on a one-dimensional approximation, the Poisson equation for the channel side is

$$\frac{d^2\phi}{dX^2} = \frac{qN_B e^{-X/L}}{\varepsilon_{si}} \tag{3.13}$$

where N_B is the epitaxial-layer doping concentration. The channel-drain junction appears at $x = 0$. The exponential-length constant L is chosen so that channel doping is maximum at the source-channel junction ($x = -X_A$); that is,

$$L = \frac{X_A}{\ln(N_{max}/N_B)} \tag{3.14}$$

where N_{max} is the maximum channel doping concentration. The electric field at $x = 0$ can be obtained by integrating the charge distribution once,

$$E_{x=0} = \frac{q}{\varepsilon_{si}} N_B L \left(e^{X_A/L} - 1 \right) \tag{3.15}$$

The voltage drop across the depletion layer of the channel side ϕ_c is obtained by integrating the electric-field distribution,

$$\phi_c = \frac{q}{\varepsilon_{si}} N_B L^2 \left[(1 + X_A/L)\, e^{X_A/L} - 1 \right] \tag{3.16}$$

The voltage drop across the depletion layer of the epitaxial N^- side ϕ_d before reachthrough can be similarly calculated by integrating

$$\frac{d^2\phi}{dX^2} = \frac{qN_B}{\varepsilon_{si}} \tag{3.17}$$

twice, which results in

$$\phi_d = \int_0^{X_d} \frac{qN_B X}{\varepsilon_{si}}\, dX = \frac{qN_B X_d^2}{2\varepsilon_{si}} \tag{3.18}$$

where X_d is the depletion-layer width of the N^- side and can be solved from the charge-neutrality relation when punchthrough occurs,

$$X_d = L\left(e^{X_A/L} - 1\right) \tag{3.19}$$

which is valid for $X_d < W_B$ (epitaxial-layer thickness). The voltage drop across the N^- side at punchthrough is obtained by substituting this expression into Eq. (3.18). Overall punchthrough voltage then becomes

$$V_{PT} = \phi_c + \phi_d \tag{3.20}$$

If punchthrough occurs after the epitaxial layer is completely depleted (reachthrough), the voltage drop across the N^- side [3.10] becomes

$$\phi_d = E_{x=0} W_B - \frac{qN_B W_B^2}{2\varepsilon_{si}} \tag{3.21}$$

which should be used in Eq. (3.20) to calculate V_{PT}.

Figure 3.6 is a plot of source-drain punchthrough breakdown voltage calculated from Eq. (3.20) with V_T and epitaxial concentrations as parameters. The threshold voltage was obtained [3.11] from

$$V_T = \phi_{ms} - \frac{Q_f}{C_o} + 2\phi_f + \frac{(4q\varepsilon_{si}N_{peak}\phi_f)^{1/2}}{C_o} \tag{3.22}$$

for a 1000 Å gate oxide. The maximum surface impurity concentration N_{peak} in the channel, used in the threshold-voltage calculation, is lower than in the substrate direction N_{max}. Depending on the processing sequence, doping method, and junction depth, the ratio of N_{peak}/N_{max} will vary; in Fig. 3.6, a ratio of 0.5 was chosen. Epitaxial thickness and doping concentration were obtained from Eq. (2.3) for a BV of 350 V.

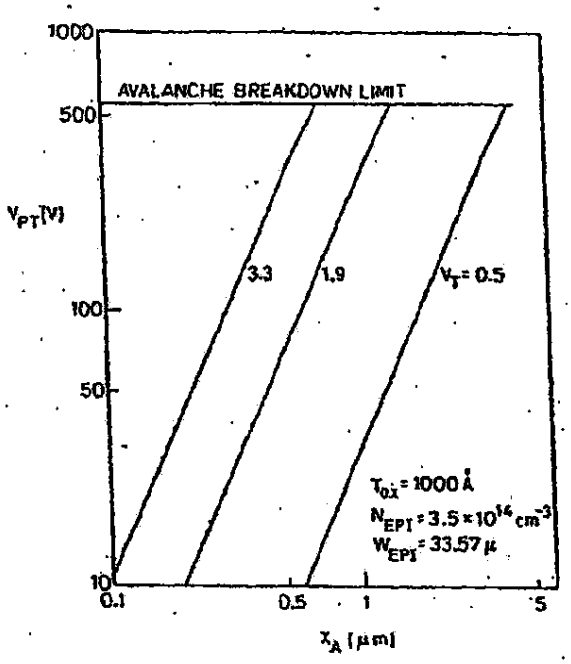


Fig. 3.6. CALCULATED SOURCE-DRAIN PUNCHTHROUGH BREAKDOWN ASSUMING AN EXPONENTIAL PROFILE.

2. Linear Approximation of the Channel Profile .

Before N^- reachthrough, Pocha's calculation [3.9] of punch-through voltage can be applied directly to the vertical structures as

$$V_{PT} = \frac{q}{\varepsilon_{si}} \left(\frac{N_{max}}{6} + \frac{N_{max}^2}{8N_B} \right) X_A^2 \qquad (3.23)$$

After N^- reachthrough, V_{PT} can be determined by solving the one-dimensional Poisson equation (similar to the exponential approximation), which results in

$$V_{PT} = \frac{qN_{max}X_A^2}{6\varepsilon_{si}} + \frac{qN_B}{2\varepsilon_{si}} (N_{max}X_A - N_B W_B) \qquad (3.24)$$

The plot in Fig. 3.7 is similar to Fig. 3.6, with the same epitaxial thickness and doping concentration and oxide thickness. For comparison, the results calculated from a step approximation of the channel profile are also plotted.

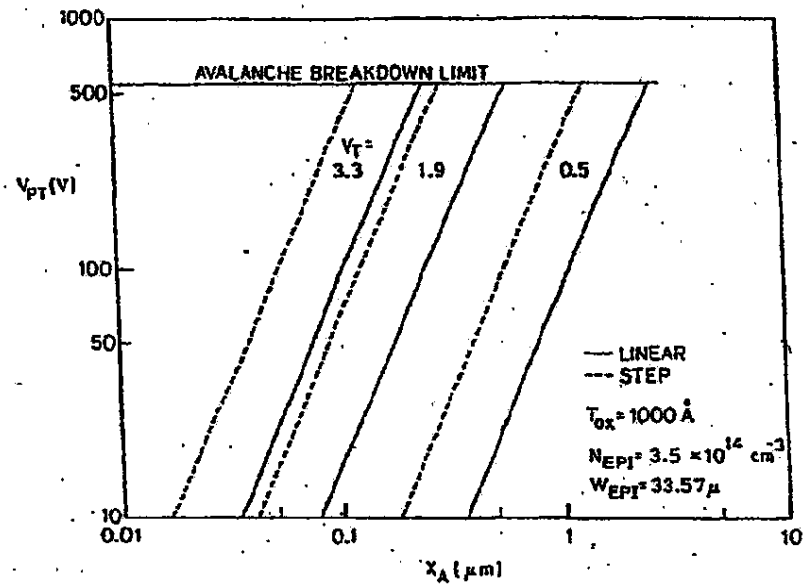


Fig. 3.7. CALCULATED SOURCE-DRAIN PUNCHTHROUGH BREAKDOWN ASSUMING LINEAR AND STEP PROFILES.

From the intersected points of avalanche and punchthrough breakdown in these two figures, minimum channel thickness $X_{A(min)}$ is plotted in Fig. 3.8 as a function of threshold voltage (which is directly related to peak channel doping). Based on Fig. 3.8, the following observations become immediately apparent.

- The exponential function yields the highest $X_{A(min)}$ for a given V_T, and the step function produces the lowest. The exponential profile is too conservative, however, because it underestimates the channel impurity dopant; on the other hand, the step profile overestimates the dopant. This is demonstrated in Fig. 3.9 where it can be seen that the linear approximation closely resembles the SUPREM simulated

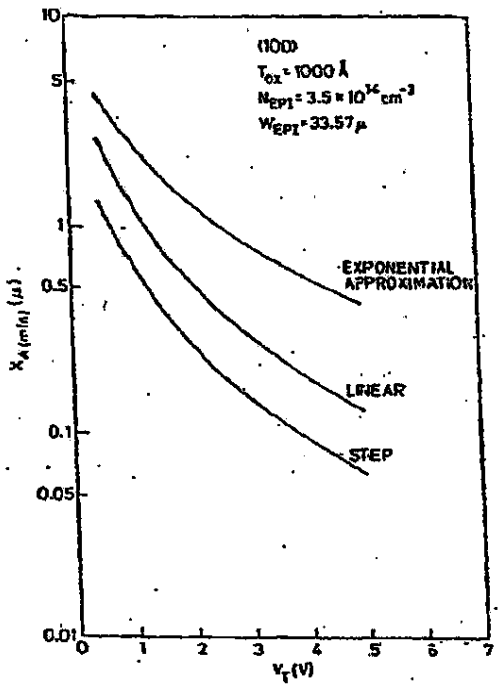
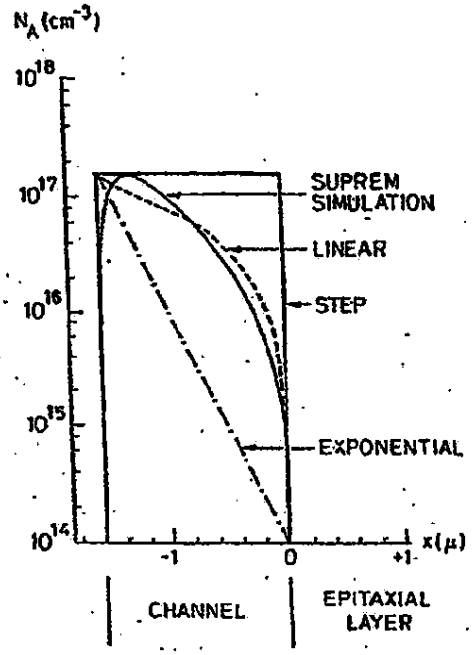


Fig. 3.8. MINIMUM CHANNEL THICKNESS ALLOWABLE TO AVOID PUNCHTHROUGH BREAKDOWN AS A FUNCTION OF THRESHOLD VOLTAGE.

profile. In addition, the resulting $X_{A(min)}$ calculated from the exponential and step functions should set the upper and lower bounds on $X_{A(min)}$ for a specified V_T.

For a specific breakdown requirement, the epitaxial-layer thickness and doping concentration can be determined as in Chapter II.E. After these parameters are known, the minimum total impurity density Q_B in the channel for a given distribution $N_C(x)$ can be obtained to avoid punch-through breakdown before avalanche occurs. Here, Q_B determines the trade-offs between V_T and channel length L_{eff} because Q_B is related to $N_C(x)$ and L_{eff} through

$$Q_B \cong q \int_0^{L_{eff}} N_C(x)\, dx \qquad (3.25)$$



step approximation = upper limit
exponential approximation = lower limit

Fig. 3.9. COMPARISON OF THREE APPROXIMATED CHANNEL PROFILES.

A shorter channel length implies a greater N_{max} and higher V_T; however, a lower V_T sets a lower bound on L_{eff}.

- In power MOSFET design, V_T is generally between 2 and 4 V. If it is too low, the noise margin will be poor and the device may not be turned off at high temperatures because of the negative temperature coefficient of V_T. If it is too high, the device cannot be driven directly by low-voltage logic circuits. After V_T is set, therefore, the punchthrough limitation will place a lower bound on L_{eff}.

C. Junction-Edge Avalanche-Breakdown Limit

The ideal breakdown voltage of a power MOSFET is bulk avalanche breakdown which corresponds to a minimum epitaxial bulk-resistance requirement (Chapter II.D). At high voltages, however, the maximum drain potential is limited by junction-edge breakdown below the ideal value because of the effects of curvature and surface electric-field crowding. To fulfill a certain breakdown requirement, therefore, epitaxial-layer doping and thickness must exceed those specified by the minimum epitaxial bulk-resistance requirement, and this will increase device on-resistance.

There are many approaches to minimizing the surface effects and to increasing junction curvature so as to attain the ideal bulk breakdown. This section reviews some of these edge-termination techniques for the VDMOS and VMOS, and those for raising the LDMOS breakdown voltage will be discussed in Chapter V.

1. VDMOS Edge Termination

The techniques discussed here are compatible with planar technology and, as a result, they can be incorporated in VDMOS fabrication with few or no additional steps. The bevelling and mesa methods are not considered because of their incompatibility with MOS processing.

a. Floating Diffused Guard Rings

Kao and Wolley [3.12] reported that concentric-ring junctions on the surface of substrates can prevent surface breakdown of the planar junctions. Figure 3.10 is a cross section of these floating guard rings. As V_{DS} increases, the source depletion region reaches the first of these "floating" P-N^- junctions, and this reachthrough establishes a self-bias which is approximately a linear function of V_{DS}. This process is then repeated at a higher drain voltage for each succeeding junction. These guard rings enlarge the effective radius of curvature by spreading the potential drop across a wide depletion region along the surface at the edge of the device. It is desirable to spread the lateral depletion over two to three times the epitaxial thickness.

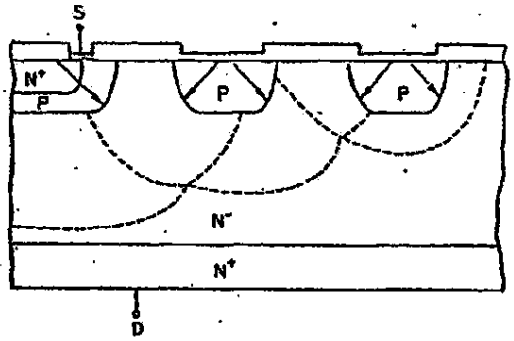


Fig. 3.10. CROSS SECTION OF THE CONCENTRIC GUARD RINGS FOR THE VDMOS.

This type of edge termination was used in the study of power MOSFETs, and Fig. 2.4b is a photomicrograph of a VDMOS with guard rings. Several commercial VDMOS devices [3.13] have adopted this technique. With the correct spacing and reasonable number of rings, greater than 80 percent of the plane breakdown value is achievable.

b. Junction-Termination Extension with a Field Plate

Figure 3.11 is a cross section of a VDMOS equipped with a metal or highly doped polysilicon field plate and equipotential (annular) rings [3.14,3.15,3.16]. The field plate is connected to the diffused junction and surrounded by a metal ring in contact with the substrate. The edge of the field plate on the oxide is much like an extension of the junction in that it lowers the junction curvature field and improves the stability of the surface potential because of a uniform potential distribution between the field plate and equipotential ring. As with the floating diffused guard rings, lateral depletion over two to three times the epitaxial thickness is again desirable, which implies that the lateral distance of the overlay of the field plate and the spacing between it and the annular ring should be approximately three times the vertical distance between the P-N^- junction and N^--N^+ interface.

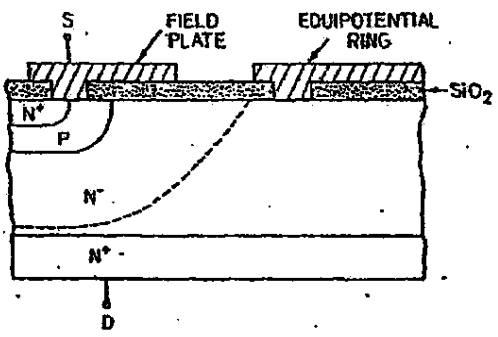


Fig. 3.11. CROSS SECTION OF A VDMOS WITH A FIELD PLATE AND AN EQUIPOTENTIAL RING.

This technique has been applied in the industrial fabrication of high-voltage bipolar transistors [3.17] and power rectifiers [3.18].

c. Junction-Termination Extension with an Implant

The concept of ion implantation to control the shape and extent of the depletion layer on the edge of the junction is similar to an ion-implanted offset gate [3.19] where the depleted offset region results in a remarkable reduction of the drain electric field. This technique has been used extensively in planar high-voltage integrated circuits [3.20] and can also be applied to the vertical structures. In Fig. 3.12 [3.21], the diffused-channel p-region is extended by a light boron implant containing a total dose of 0.5 to 0.9 $\epsilon_{si} E_{crit}/q$, depending on the surface-passivant quality (typically 1×10^{12} ions/cm^2). The lightly implanted P^--N^-

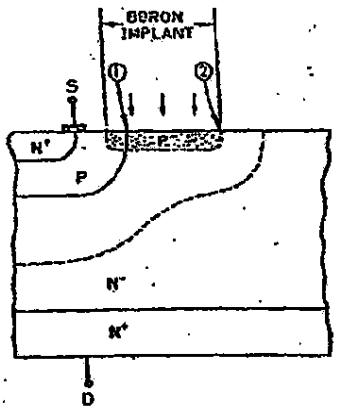


Fig. 3.12. CROSS SECTION OF A VDMOS WITH AN EDGE-EXTENSION IMPLANT.

junction results in a low peak surface field at (1) and (3); the higher the implant dose, the lower the peak surface field at (2) and the larger the peak field at (1). On the other hand, the heavier the implant dose, the higher the field at (2). The experimental data (approximately 75 to 80 percent bulk value) are plotted in Fig. 3.13 as are the experimental results obtained from this and other studies of the concentric guard rings [3.12,3.13,3.22,3.23] and the field plate and annular ring [3.17, 3.18,3.23].

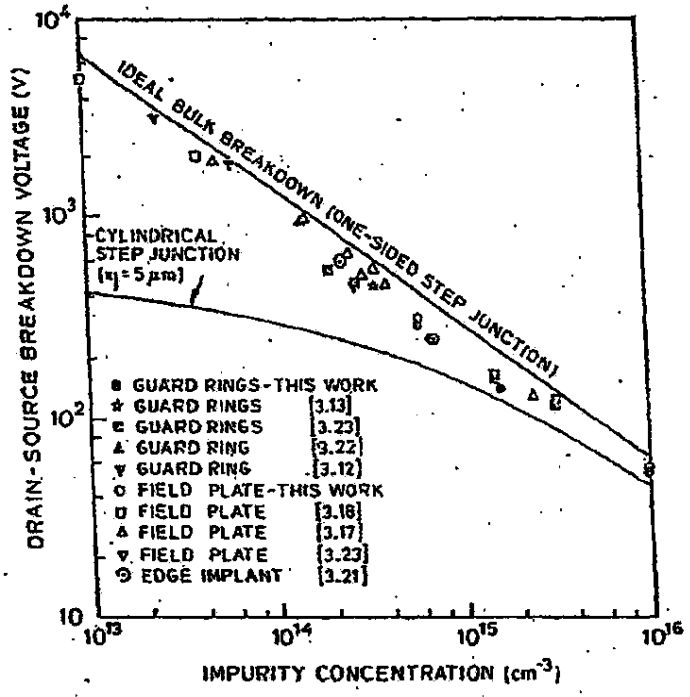


Fig. 3.13. BREAKDOWN-VOLTAGE MEASUREMENTS TO COMPARE THE EDGE-TERMINATION TECHNIQUES FOR VDMOS DEVICES.

2. VMOS Edge Termination

The methods employed to increase VMOS breakdown voltage are similar to the guard rings, field plate, and edge implant for the VDMOS.

There are variations in the geometry, however, because the VMOS is a nonplanar device.

a. V-Groove Floating Guard Ring [3.24]

By utilizing concentric grooves around the device periphery, isolated floating guard rings can be fabricated while the V-groove channels are being etched, as shown in Fig. 3.14. The floating-ring spacing at the corners is very difficult to control, however, because of the anisotropic etching of various crystalline orientations around the corners. Experimental results are roughly 80 percent of the plane limit.

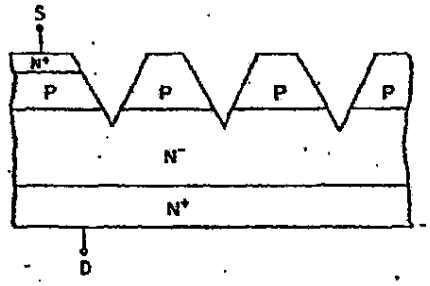


Fig. 3.14. V-GROOVE FIELD-LIMITING RING.

b. Junction-Termination Extension with an Implant [3.21]

The junction can be terminated by the same etch that forms the channels (Fig. 3.15). The p-region at the junction is then extended by a light boron implant, similar to that of the VDMOS. (Edge termination by etching alone without an implant is called "plain mesa.") Experimental results ranged from 75 to 90 percent of the ideal values.

c. Junction-Termination Extension with a Field Plate [3.25]

In this geometry, the field plate is placed over the etched junction (Fig. 3.16) and, when its potential has the correct polarity with respect to the bulk silicon, the space charge of the depleted

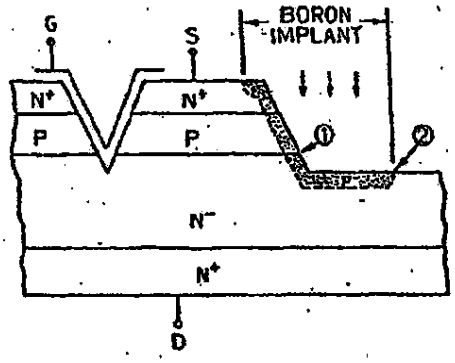


Fig. 3.15. JUNCTION-TERMINATION EXTENSION WITH AN IMPLANT.

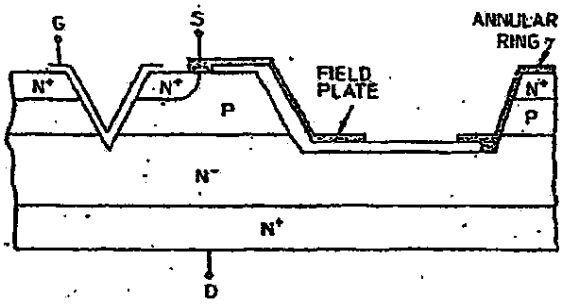


Fig. 3.16. JUNCTION-TERMINATION EXTENSION WITH A FIELD PLATE.

layer in the epitaxial region is balanced by the charge on the field plate. The equipotential lines of this nonplanar field plate are expected to be smoother at the pn^- metallurgical junction than those of the planar field plate. The experimental result of 200 V on 5 Ω-cm material is lower than obtained from the guard rings but higher than the plain-mesa type, as can be seen in Fig. 3.17. Also plotted in this figure for comparison are the results obtained by using three diffused guard rings (developed in this study), the V-groove floating guard ring [3.24], and the junction-edge extension with an implant [3.21].

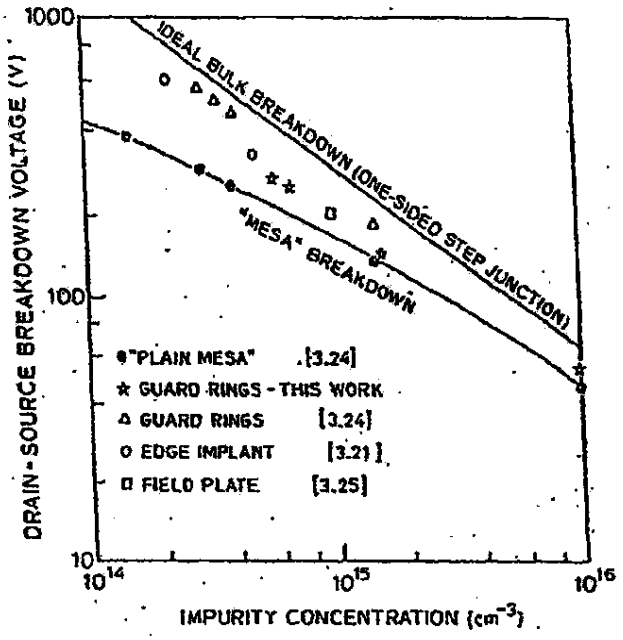


Fig. 3.17. BREAKDOWN-VOLTAGE MEASUREMENTS TO COMPARE THE EDGE-TERMINATION TECHNIQUES FOR VMOS DEVICES.

3. Comparison of Junction-Termination Techniques

For a fair comparison of the above techniques, the termination structure must be optimized. For example, when using guard rings, their number, the spacing between them, and the width of each are critical factors; when using a field plate, oxide thickness, the field-plate overlapping distance, and the spacing between it and the annular ring must be considered. Such second-order effects as breakdown stability must also be taken into account with respect to long-term reliability.

The choice of method depends on

- efficient utilization of the silicon area to meet the breakdown requirement (this requires a trade-off between on-resistance and breakdown voltage)

compatibility with a given fabrication process (for example, guard rings are impractical in low-voltage devices because of the small depletion-layer width and, as a result, the field plate would be a better choice; in high-voltage devices, the guard ring is less sensitive to such processing constraints as oxide thickness and metal-polysilicon crossover)

D. Parasitic Bipolar Latchback Limit

A complete analysis of the parasitic bipolar transistor in power MOSFETs is an enormous task because of the three-dimensional nature of the problem; not only are the device vertical profiles involved, but layout topology also plays a key role in the parasitics. As a result, a one-dimensional approximation is used in this section to analyze the influence of the vertical profiles on the static and dynamic latchback behavior of the bipolar transistor.

1. Static Condition

Because of the double diffusions (or even diffusion in the epitaxial channel) in typical power MOSFETs, the region under the N^+ source becomes the pinched resistor (analogous to the pinched base of the bipolar transistor) and the region under the gate is similar to a bipolar transistor with its base shorted to the emitter through the pinched resistor R_p (see Fig. 3.16). The R_p value can be hundreds of ohms to tens of kilohms and can rise even higher under strong drain-to-channel bias. Because the channel contact is not adjacent to the parasitic bipolar transistor under the gate, this resistance is crucial in determining the maximum operating voltage. The BV_{CEO} of the bipolar transistor is considerably less than the intrinsic value of the breakdown voltage in the drain-channel junction BV_{CBO} and, as a result, the minimum channel thickness set by the punchthrough limit could be reduced to a second-order consideration by the parasitic bipolar latchback limit.

The breakdown behavior can be explained by a simple one-dimensional model based on the parasitic bipolar transistor in parallel with the intrinsic MOS (Fig. 3.16b). Turn-on of this parasitic can be the result of

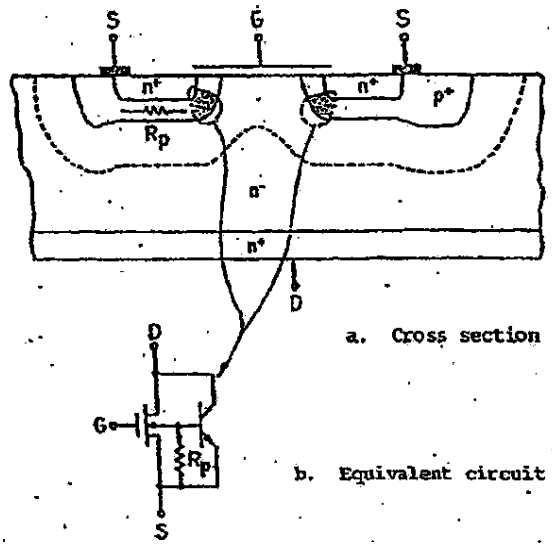


Fig. 3.18. PARASITIC BIPOLAR TRANSISTOR INTRINSIC TO THE DMOS STRUCTURE.

- an increase in the reverse leakage current I_{CBO} and pinched channel resistance with applied drain bias
- low-level carrier multiplication [3.26] within the drain-channel space-charge layer in the vicinity of the region immediately under the gate where the electric field is generally the highest

The avalanche hole current flows from the channel near or under the gate to the channel contact through R_p. This produces the voltage drop across R_p which, at the onset of breakdown, forward biases the emitter-base junction of the parasitic. MOS drain-source breakdown is equivalent, therefore, to collector-emitter breakdown BV_{CEO} in a bipolar transistor.

The current flowing through R_p [3.27] is approximately

$$I_B = MI_{CBO} - (M - 1) I_E + \frac{I_E}{1 + \beta} \qquad (3.26)$$

where

M = low-level multiplication factor

I_{CBO} = drain-channel reverse leakage

β = bipolar current gain

I_E = emitter (source) current

At avalanche breakdown BV_{CBO}, $M = \infty$; at collector-to-emitter (open base) breakdown BV_{CEO}, $\alpha_o M = 1$, where M is only slightly greater than 1 and α_o is the product of emitter efficiency and the base-transport factor. By reducing the two-dimensional effect to a one-dimensional formulation, the voltage drop across R_p becomes

$$V_{BE} = \frac{1}{3} R_p I_B \tag{3.27}$$

where the factor of 1/3 is based on the same geometrical consideration as discussed in Chapter II.E and illustrated in Fig. 3.19.

Before the bipolar device is forward biased, emitter current I_E is very small. Only the first term in Eq. (3.26) is significant and,

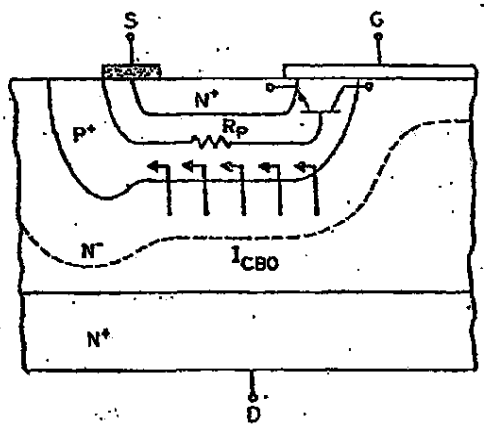


Fig. 3.19. CROSS SECTION OF CURRENT FLOW I_{CBO} UNDER REVERSE BIAS.

therefore, Eq. (3.26) becomes $I_B \approx MI_{CBO}$. In the useful range of weak avalanche in the bipolar transistor where $M < 1.1$ [3.28], I_B can be considered as

$$I_{CBO} = \frac{Aqn_i}{2\tau_{sc}} X_{dn} \qquad (3.28)$$

with negligible error [3:29], where

A = effective channel-drain junction area

n_i = intrinsic carrier concentration

τ_{sc} = space-charge generation lifetime

X_{dn} = depletion-layer width on the channel side

Channel resistance R_p is equal to sheet resistance $R_\square$ multiplied by the number of squares under the source. Here, $R_\square$ is defined as

$$R_\square = \left(\int_0^{W_c} qN_c(x)\, \mu(N_c)\, dx \right)^{-1} \qquad (3.29)$$

where

$N_c(x)$ = channel doping concentration

$\mu(N_c)$ = hole bulk mobility whose doping dependence is reported in Ref. 3.36

W_c = net channel thickness (V_{DS}-dependent)

It should be noted that both I_{CBO} and R_p are voltage-dependent.

Bipolar turn-on will occur when reverse bias V_{DS} causes V_{BE} to be ≈0.6 V. The initiation of latchback may not be necessarily at $V_{DS} = BV_{CBO}$ but it can occur at a lower V_{DS}, depending on device geometry and fabrication process. Under the latchback condition, device breakdown voltage BV_{DSS} is related empirically to BV_{CBO} [3.31] by

$$BV_{DSS} \approx BV_{CEO} \approx \frac{BV_{CBO}}{\beta^{1/n}} \qquad (3.30)$$

where $4 < n < 7$ for $50 < BV_{CBO} < 1000$ V [3.28]. Relating β to X_A and minority-carrier lifetime τ_n, BV_{CEO} [3.27] becomes

$$BV_{CEO} \cong BV_{CBO}\left(\frac{X_A^2}{2D_n\tau_n}\right)^{1/n} \tag{3.31}$$

where D_n is electron diffusivity.

After device layout topology is established, this first-order theory will enable the calculation of minimum channel thickness when bipolar latchback occurs. For example, the linear-profile approximation is used because it is close to the simulated profile, and Fig. 3.20 will facilitate the V_{BE} calculation. To simplify the analysis, the effective resistive length d_{eff} is assumed to include the two-dimensional edge effects. Substituting Eq. (3.28) into (3.27) yields

$$V_{BE} = \frac{1}{3} R_D d_{eff}^2 \frac{qn_i}{2\tau_{sc}} X_{dn} \tag{3.32}$$

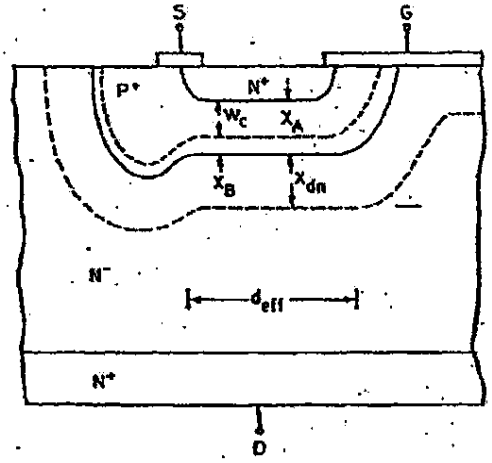


Fig. 3.20. DEVICE CROSS SECTION USED TO CALCULATE MINIMUM CHANNEL THICKNESS.

and channel sheet resistance as derived in Eq. (3.29) becomes

$$R_{\square} = \frac{1}{q\left(X_A - X_B^2/X_A\right)\bar{\mu}N_{max}/2} \tag{3.33}$$

where

N_{max} = peak channel doping concentration

X_A = channel thickness

X_B = depletion-layer width on the channel side

$\bar{\mu}$ = average hole mobility in the channel

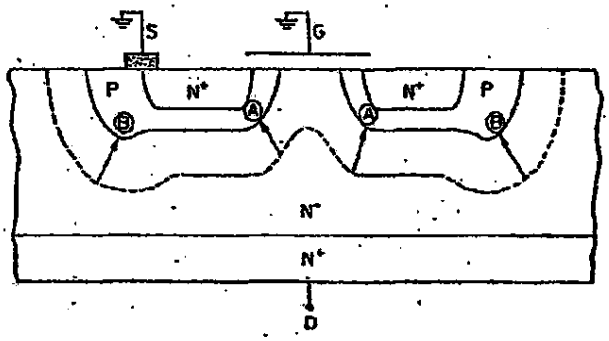
The depletion-layer width on the channel side X_B for a given V_{DS} can be obtained from a similar calculation as in Section B.2 and is the solution to

$$V_{DS} + \phi_B = \frac{qN_{max}X_B^3}{6\epsilon_{si}X_A} + \frac{qN_D X_{dn}}{2\epsilon_{si}} \tag{3.34a}$$

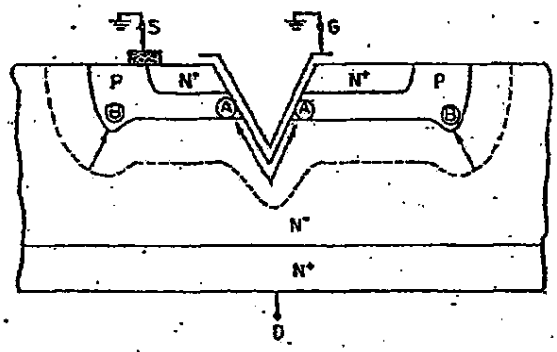
$$X_{dn} = \frac{N_{max}X_B^2}{2N_D X_A} \tag{3.34b}$$

Bipolar turn-on is most likely to occur when V_{DS} is close to the breakdown voltage, where I_{CBO} and $R_{\square}$ approach their highest values. Based on the same epitaxial thickness and resistivity as in Section B and on τ_{sc} = 10 μsec, X_A [derived in Eq. (3.32) and with V_{BE} = 0.6 V] is slightly higher than the value obtained from the punchthrough breakdown limit (the difference is $<10^{-3}$ μ); therefore, the $X_{A(min)}$ of static bipolar turn-on is essentially the same as the $X_{A(min)}$ of punchthrough breakdown.

The above analysis is one dimensional. If the curvature effect is taken into account, such as device avalanche at Ⓐ in Fig. 3.21, latchback will still occur even in devices with greater channel thickness because of the very large avalanching hole current passing through $R_{\square}$. Avalanche breakdown should occur, therefore, as close to the channel



a. VDMOS



b. VMOS

Fig. 3.21. CROSS SECTION OF HIGH ELECTRIC-FIELD POSITIONS.

contact as possible, such as (B) in Fig. 3.21; in properly designed VDMOS and VMOS devices, this is exactly what takes place. Merging of the depletion regions under the gate in the VDMOS reduces channel-junction curvature and, similarly, the correct V-groove depth minimizes the field strength at the apex. Avalanche will occur at the edge of the device.

One of the devices having the latchback phenomenon is shown in Fig. 3.22 [3.32]. Avalanche probably occurs at the apex of the overetched V-groove, which forces the hole current to flow through R_B and results in the turn-on of the parasitic.

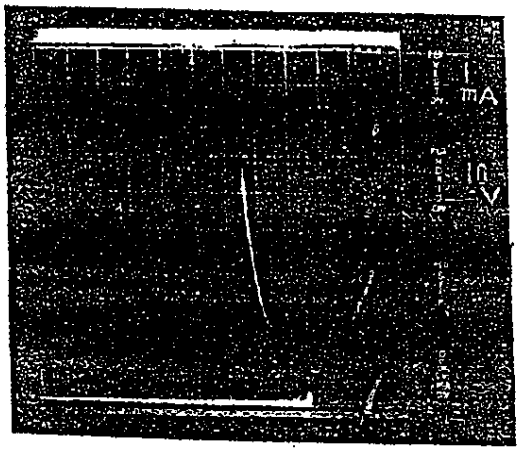


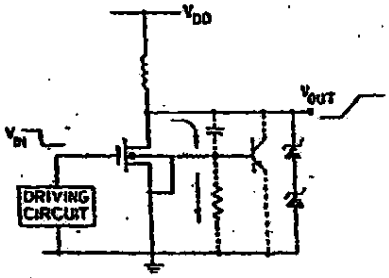
Fig. 3.22. I-V CHARACTERISTICS OF THE POWER MOSFET IN REVERSE BREAKDOWN.

2. Dynamic Condition--Limitation on the Switching Transient dV/dt

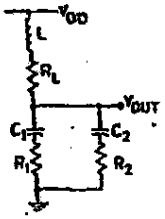
Under certain switching conditions and loads, the parasitic bipolar NPN can be turned on by the large capacitive current flowing through the pinched-channel resistor. In the example in Fig. 3.23a, the power MOSFET has an inductive load and the output is connected to zener diodes. When the MOS is suddenly turned off, the output voltage will rise from $V_{DS(SAT)}$ at a rate of dV/dt which can be determined from the equivalent circuit in Fig. 3.23b. Because the channel-drain junction changes from small to high reverse bias in a short time, the large amount of capacitive current that results from the variation in the channel-drain space-charge region can potentially forward bias a portion of the p-well region (Fig. 3.23c). The injected charges stored in the channel could make the turn-off characteristics appear similar to those of the bipolar device.

Based on the one-sided abrupt-junction approximation, the depletion-layer charge on the N^- drain side is

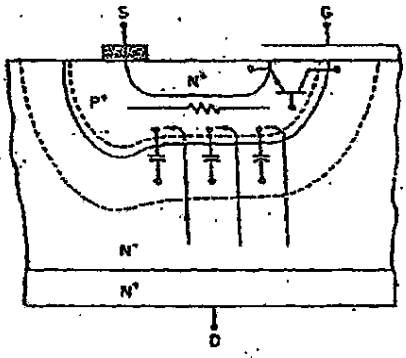
$$Q_D = \left[2q\varepsilon_{si}N_D(V_{DS} + \phi_B)\right]^{1/2} \qquad (3.35)$$



a. Schematic



b. Equivalent circuit during turn-off



c. Cross section

Fig. 3.23. DEVICE TURN-OFF RESPONSE.

FCS1689341

FCS1692840

and the switching current density is

$$J_D = \frac{dQ_D}{dt} = \left[\frac{q\epsilon_{si}N_D}{2(V_{DS} + \phi_B)}\right]^{1/2} \frac{dV_{DS}}{dt} \qquad (3.36)$$

The source-channel junction voltage, V_{BE} can be obtained from the product of J_D [Eq. (3.36)] and R_p [Eq. (3.29)]. If V_{BE} is ≈0.6 V during the turn-off transient, the injected minority carriers will slow down dV/dt and thereby prolong the turn-off time. It should be noted that even the final V_{DS} (and the overshoot) may not be sufficiently high to cause low-level avalanche-induced parasitic turn-on; the fast-changing dV/dt current could turn on the parasitic BJT.

Parasitic bipolar turn-on caused by dV/dt is illustrated in Fig. 3.24 [3.33]. Turn-on occurs at $V_{DS} \approx 125$ V because of the steep initial slope of dV/dt, and the injected carriers stored in the channel then slow down dV/dt and produce a current spike. After the recombination of excessive carriers in the channel, the voltage again begins to

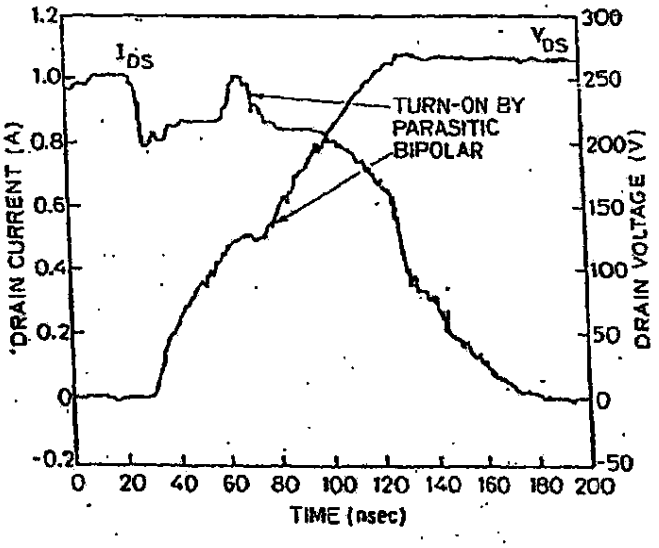


Fig. 3.24. CURRENT AND VOLTAGE WAVEFORMS INDICATING TURN-ON OF THE PARASITIC BIPOLAR TRANSISTOR IN A 300 V EXPERIMENTAL GENERAL ELECTRIC POWER MOSFET [3.33].

rise. The chance for bipolar turn-on is less as V_{DS} becomes larger, and this is apparent in Eq. (3.36).

The following observations are the result of examination of Fig. 3.24.

- Bipolar turn-on increases the power MOSFET turn-off time, and this additional delay is approximately 30 nsec (23 percent longer).
- Fortunately, bipolar turn-on occurred at $V_{DS} = 125$ V which is below the bipolar BV_{CEO} breakdown voltage; the MOS drain voltage was able to rise again after the excess carriers were recombined. If the turn-on caused by dV/dt had occurred at $V_{DS} \geq BV_{CEO}$, the output drain voltage would have been clamped and would have remained at BV_{CEO} provided that the load could supply current greater than the triggering current I_{CEO}.

The limitation imposed by the dV/dt requirement on $X_{A(min)}$ (minimum channel thickness as determined in Section 2.B) could exceed that imposed by punchthrough breakdown. To apply the above theory to the VDMOS, the same epitaxial doping concentration as in the punchthrough breakdown limit ($N_D = 3.5 \times 10^{14}$ cm^{-3}) is used, and the threshold-voltage calculation is identical to the previous formulations. As in Section 1 above, Fig. 3.20 serves as the basis for the V_{BE} derivation.

Based on the linear-profile approximation, V_{BE} becomes

$$V_{BE} = \frac{1}{3} R_D d_{eff}^2 J_d \qquad (3.37)$$

where R_D and J_d are defined in Eqs. (3.33) and (3.36), respectively. The experimentally observed dV/dt is between 100 [3.13] and 5 [3.33] V/nsec; 100 V/nsec was chosen for these calculations. The minimum channel thickness $X_{A(min)}$ is set by the parasitic bipolar turn-on caused by fast dV/dt and is calculated from Eq. (3.37) with $V_{BE} = 0.6$ V and V_{DS} ranging from $V_{DS(SAT)}$ (≈0.5 V) to BV_{DSS} (≈450 V). The results thus obtained at $V_G = 0.5$ and 1.9 V are plotted in Fig. 3.25. For $d_{eff} < 6$ μ, $X_{A(min)}$ is determined by the punchthrough breakdown limit (note that $X_{A(min)}$ resulting from static bipolar turn-on is

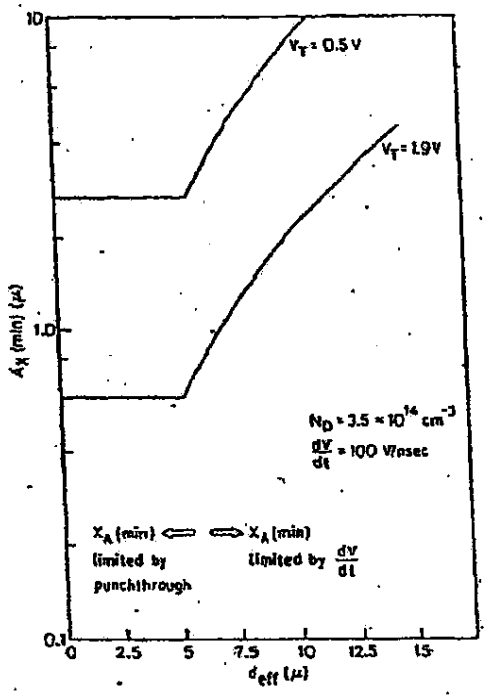


Fig. 3.25. MINIMUM CHANNEL THICKNESS AS A FUNCTION OF EFFECTIVE RESISTIVE LENGTH AT TWO THRESHOLD VOLTAGES.

approximately the same as from punchthrough). For $d_{eff} > 6$ μ, $X_{A(min)}$ is determined by the dV/dt requirement and, here, dV/dt dictates the channel profiles and device layout. The punchthrough limit dominates only when the channel contacts are adjacent to the gates (d_{eff} is small, as in Fig. 3.20). Similar calculations can be performed for the VMOS devices.

E. Johnson's Limit

1. Original Model

Johnson's approach [3.1] was to determine the best possible trade-off between cutoff frequency f_T and maximum allowable applied

voltage BV for any semiconductor structure, based on the following assumptions.

- There is a maximum possible "saturated drift velocity" of carriers v_{SAT} in a semiconductor. This is a material constant on the order of 6×10^6 cm/sec for electrons in silicon.
- There is a maximum electric field E_{crit} that can be sustained in a semiconductor without dielectric breakdown. This is approximately 2×10^5 V/cm in silicon.

With these two approximations and the simplified transistor model in Fig. 3.26, the cutoff frequency and maximum applied voltage can be expressed as

$$BV = E_{crit} l \tag{3.38a}$$

$$f_T = \frac{1}{2\pi\tau_T} = \frac{1}{2\pi l/v_{SAT}} \tag{3.38b}$$

which, for silicon, yields

$$BV \cdot f_T = \frac{E_{crit} v_{SAT}}{2\pi} = 200 \text{ V-GHz} \tag{3.39}$$

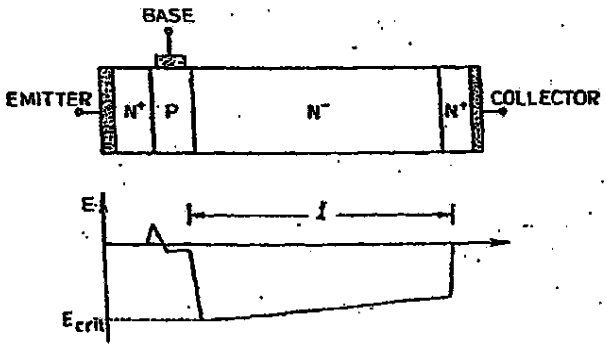


Fig. 3.26. IDEALIZED MODEL OF THE ONE-DIMENSIONAL CURRENT-FLOW TRANSISTOR.

It should be noted that this derivation is the result of the following simplifications.

- The N^- region is lightly doped so that E_{crit} remains constant across it.
- Because the P-region is relatively short in comparison to the N^-, transit time in the P-region (base of BJT, channel of FET) can be neglected.
- RC time delays caused by series resistance and junction capacitances are not included in the f_T calculation.

As a result, Eq. (3.59) is regarded as the maximum figure of merit for the voltage/frequency relationship.

2. Modified Limit

The v_{SAT} in Eq. (3.39) directly affects the magnitude of Johnson's limit. In VDMOS, VMOS, and bipolar transistors, the drift region is in the semiconductor bulk, and v_{SAT} is the electron saturated drift velocity in bulk silicon. In the LDMOS, part of the drift region is in the surface accumulation layer although the major portion is in the bulk; however, v_{SAT} in the accumulation layer is not substantially different from the bulk value.

More accurate measurements of v_{SAT} [3.34,3.35] indicate that it is approximately 1×10^7 cm/sec for an electric field higher than 5×10^4 V/cm at 300°K. Based on this data, Johnson's modified limit for silicon with a lightly doped N^- layer ($N_D \leq 1 \times 10^{14}$ cm^{-3}) becomes

$$BV \cdot f_T \leq \frac{2 \times 10^5 \times 10^7}{2\pi} = 318 \text{ V-GHz} \qquad (3.40)$$

3. Parametric Dependence

a. Temperature

Based on the experimental data, v_{SAT} as a function of temperature can be represented [3.36] by

$$v_{SAT} = \frac{v^*}{1 + C^* \exp(T/\Theta)} \qquad (3.41)$$

where $v^* = 2.4 \times 10^7$ cm/sec, $C^* = 0.8$, and $\Theta = 600°K$.

The temperature dependence of E_{crit} in a silicon abrupt junction can be calculated from the temperature dependence of the breakdown voltage [3.37] by using the following equation:

$$E_{crit} = \left(\frac{2qN_D BV}{\varepsilon_{si}}\right)^{1/2} \qquad (3.42)$$

Figure 3.27 plots E_{crit} as a function of T at three doping concentrations; it can be seen that the fractional change in E_{crit} with T decreases with increasing N_D. The temperature variation of Johnson's limit for silicon when $N_D = 1 \times 10^{14}$ cm^{-3} is shown in Fig. 3.28, based on the temperature-dependence data in Eq. (3.41) and Fig. 3.27. It should be noted that this limit is relatively insensitive to temperature and reaches its peak value at room temperature.

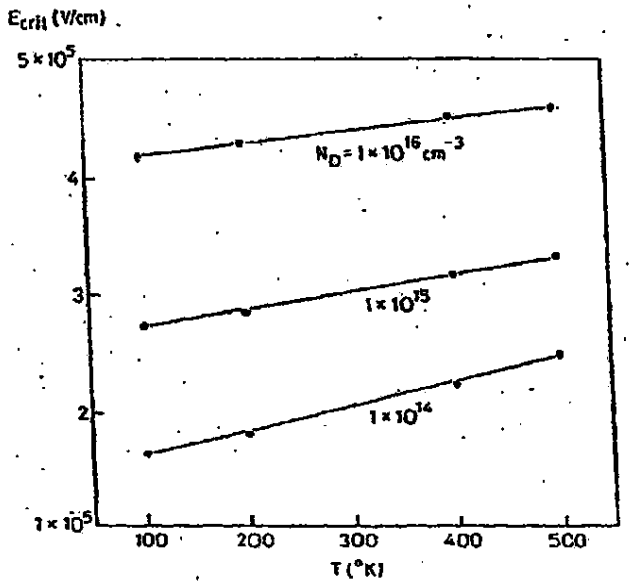


Fig. 3.27. CRITICAL FIELD FOR AN ABRUPT SILICON JUNCTION AS A FUNCTION OF TEMPERATURE.

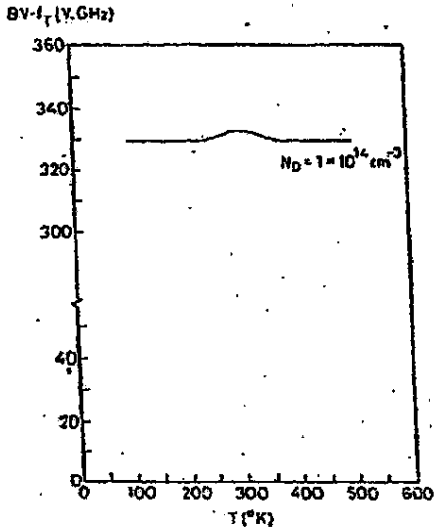


Fig. 3.28. JOHNSON'S LIMIT AS A FUNCTION OF TEMPERATURE.

b. Material Properties

The performance of power MOSFETs is ultimately influenced by the basic material properties and, as a result, Johnson's limit is also expected to be affected by them. For example, a material with a wide energy band gap generally has larger critical fields of avalanche breakdown, and a material with a direct energy band gap has higher carrier mobilities and saturation velocities.

The limit for Ge has been obtained [3.1] as

$$BV \cdot f_T \cong 100 \text{ V.GHz} \qquad (3.43)$$

and the limit for GaAs can be calculated as follows. The critical field as determined in Ref. 3.38 relates the breakdown voltage to the energy band gap and doping concentration as

$$BV = 60\left(\frac{E_G}{1.1}\right)^{3/2}\left(\frac{N_B}{10^{16}}\right)^{-3/4} \qquad (3.44)$$

Using E_c = 1.4 eV and $N_B = 1 \times 10^{14}$ cm^{-3} and based on Eqs. (3.42) and (3.44), $E_{crit} = 2.9 \times 10^5$ V/cm. The velocity vs field curve for GaAs has an overshoot at $v_{peak} = 2.17 \times 10^7$ cm/sec when $E = 3.2 \times 10$ V/cm and then levels off at $v_{SAT} \cong 1.16 \times 10^7$ cm/sec [3.39]. If an average value of 1.2×10^7 cm/sec is chosen, Johnson's limit for GaAs would become

$$BV\text{-}f_T \cong 552 \text{ V-GHz} \qquad (3.45)$$

which is consistent with other calculations [3.40] and indicates that GaAs should have lower on-resistances and greater switching efficiency in comparison to Si and Ge.

4. Comparison of Experimental Results

The measured values of the breakdown voltage and cutoff-frequency product are consistently lower than the calculated values because saturation velocity may not be attained over the entire length of the device and the electric field may not be constant over the N^- region.

The experimental data in Table 3.1 were obtained from reported maximum breakdown voltages and experimental cutoff frequencies, and these results plus Johnson's and other bipolar data [3.45,3.46] are plotted in Fig. 3.23. The new theoretical limit [Eq. (3.40)] and the experimental values of a static-induction transistor (SIT) [3.47] are also included

Table 3.1

EXPERIMENTAL BREAKDOWN VOLTAGES AND CUTOFF FREQUENCIES IN POWER MOSFETs

Device	Breakdown Voltage (V)	Cutoff Frequency (GHz)	BV-f_T (V-GHz)	Ref.
LDMOS	200	0.8	160	3.41
LDMOS	160	1.1	176	3.42
LDMOS	35	1.4	49	3.42
VDMOS	90	1.3	117	3.43
VMOS	100	1.5	150	3.44

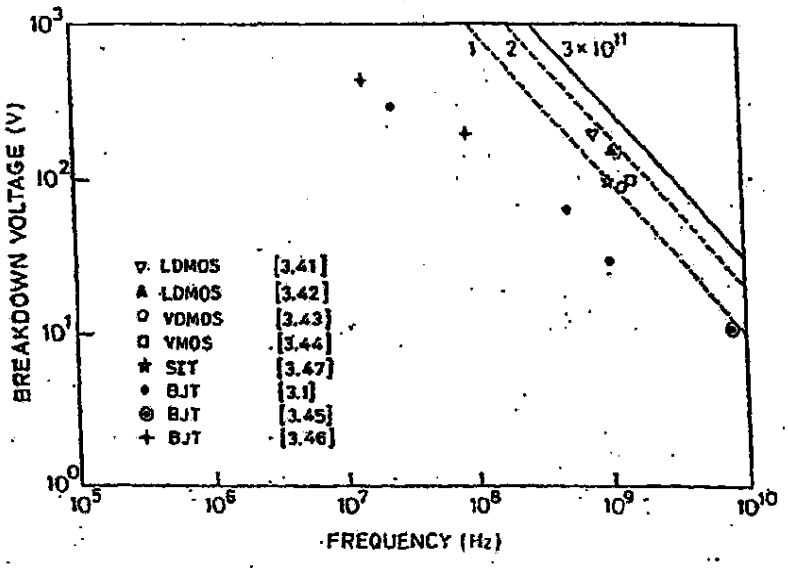


Fig. 3.29. EXPERIMENTAL VOLTAGE/FREQUENCY RELATIONSHIP IN MOS, BIPOLAR, AND JFET DEVICES.

for comparison. In the past, JFETs (of which the SIT is a short-channel version) had poor frequency responses because of their long channel lengths and large junction capacitances; however, recent technological advances enabled the realization of short channel lengths and narrow gate spacings which greatly improved their high-frequency performance.

Of the three types of devices, the MOS tends to lie closest to the theoretical curve in Fig. 3.29. In the low-voltage domain, the MOS drops away because the higher doping concentration makes the field triangular, with its peak value equal to E_{crit}, and the electron will travel only part of the distance at v_{SAT}. In Johnson's original model, however, the N^- drift region was lightly doped and, as a result, the electric field was constant across it and the electron was able to travel at saturated velocity over the entire region. In addition, the increased junction capacitances could invalidate the original assumption that f_T comes solely from the drift-region transit time. The bipolar transistor in the high-voltage domain tends to fall away from the theoretical line because of the poor f_T performance on a lightly doped substrate, as will be discussed in the following section.

5. Fundamental Differences between MOS and Bipolar Devices

The one-dimensional model in Fig. 3.30 is used to examine the differences between bipolar and MOS devices in Johnson's limit. Because the two devices are assumed to have the same vertical doping profiles, their breakdown voltages should be identical. The basic difference is in their current-gain cutoff frequency which is a measure of frequency response.

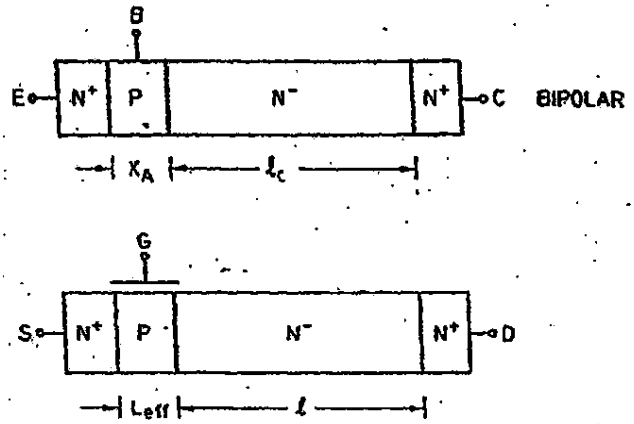


Fig. 3.30. ONE-DIMENSIONAL MODEL FOR BIPOLAR AND MOS TRANSISTORS.

The intrinsic MOS cutoff frequency (3.43) is

$$f_T = \frac{g_{m(max)}}{2\pi C_{in}} \tag{3.46}$$

where C_{in} is the input capacitance which is equal to $C_o WL_{eff}$. At the saturation-velocity limit, $g_{m(max)}$ becomes

$$g_{m(max)} = C_o W v_{SAT} \tag{3.47}$$

By substituting $g_{m(max)}$ and C_{in} into Eq. (3.46), the maximum f_T for a given channel length becomes

$$f_T = \frac{v_{SAT}}{2\pi b_{eff}} \qquad (3.48)$$

The maximum f_T of a real device will be lower than predicted by Eq. (3.48) because of the additional parasitic RC time delays and the transit time in the space-charge region between the channel and drain.

The bipolar f_T in a common-emitter configuration is expressed [3.49] as

$$\frac{1}{2\pi f_T} = r_E C_E + \frac{x_A^2}{n_b D_n} + \frac{l_c}{2v_{SAT}} + (r_E + r_C)\, C_{BC} \qquad (3.49)$$

where

r_E = emitter resistance

C_E = emitter-base capacitance

x_A = basewidth

n_b = constant (2 ~ 4)

D_n = electron diffusion constant in the bipolar base

l_c = collector depth

r_C = collector resistance

C_{BC} = base-collector capacitance

The last two terms in Eq. (3.49) are approximately the same as those for the MOS devices. The f_T difference between the bipolar and MOS is in the first two terms, where $r_E C_E$ is the emitter delay and $x_A^2/n_b D_n$ is the base transit time determined by the diffusion of minority carriers (electrons) through the base rather than by the drift of majority carriers (electrons) across the channel.

Emitter delay τ_E contains an emitter-base transition capacitance C_{TE} and a neutral capacitance C_{NE} [3.49,3.50,3.51] that comes from the additional charge in the emitter space-charge layer over and above the charge stored in C_{TE}. This C_{NE} is inversely proportional to the emitter-base concentration gradient and has been used successfully

to explain the difference in f_T between phosphorus- and arsenic-emitter bipolar transistors [3.49,3.50]; it could also constitute more than 70 percent of τ_E [3.50] in high-frequency transistors. In high-speed ECL bipolar transistors, τ_E is the dominant factor in total time delay [3.52]; however, at lower frequencies, it could become less important.

Emitter delay τ_E [3.50] is

$$\tau_E = \frac{kT}{qI_E} C_{TE} + 2G \frac{\bar{U}}{a} \qquad (3.50)$$

where

G = Gummel number [3.53]

$\bar{U}$ = normalized potential drop between the metallurgical junction and boundary of the emitter space-charge layer averaged over the range of currents considered (see Ref. 3.51)

a = emitter-base junction impurity gradient

Base transit time τ_B at high current levels contains two terms [3.54],

$$\tau_B \approx \frac{X_A^2}{4D_n} + \frac{W_{CIB}^2}{4D_n} \qquad (3.51)$$

where X_A is the quasi-neutral basewidth and W_{CIB} is the current-induced basewidth resulting from high-level injection [3.54]. In a lightly doped epitaxial layer, W_{CIB} could be several times larger than W_B; for example, in Ref. 3.54, W_{CIB}/W_B is approximately 4 near $f_{T(max)}$ for a 5.5×10^{14} cm^{-3} epitaxial-layer doping concentration and 9.4 μ epitaxial width.

In comparing MOS and bipolar delay times, the basewidth (channel length) is assumed to be _ μm and base (channel) doping is 5×10^{16} cm^{-3}. The MOS and bipolar base transit times are ≈15 and 625 psec, respectively. Emitter delay is ≈30 to 100 psec, depending on the gradients a and I_E. From these data, the following conclusions can be reached.